2017

Hate Speech Laws, Legitimacy, and Precaution: A Reply to James Weinstein

Alexander Brown

HATE SPEECH LAWS, LEGITIMACY, AND PRECAUTION: A REPLY TO JAMES WEINSTEIN

*Alexander Brown**

Ronald Dworkin once remarked to me that he thought Robert Nozick was a highly skilled defender of the indefensible. I have the impression from reading James Weinstein's interesting article that this is partly how he sees defenders of hate speech bans.[1] This is not how I see myself, of course; which is to say, I see myself as neither especially skilful nor as defending the indefensible. Indeed, given that, as I attempted to show in my recent book,[2] not only does virtually every person on the planet live under at least some form of hate speech law but also such law is marked by great internal variety, I rather suspect that what is indefensible is either rejecting or defending hate speech law en masse. I hope to bring this out in my contribution to this symposium.[3]

There is much in Weinstein's article to contemplate, but I shall limit myself to making the following four main points. First, I believe that debates concerning the normative standing of hate speech law are always improved by heeding the internal variety of

* Reader in Political and Legal Theory, University of East Anglia (UEA), UK. I am tremendously grateful to Professor Weinstein for inviting me to join this special edition and for numerous constructive and candid exchanges during the course of writing my reply.

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

2. ALEXANDER BROWN, HATE SPEECH LAW: A PHILOSOPHICAL EXAMINATION (2015).

3. I shall not, however, attempt a detailed analysis of the concept of hate speech, nor comment on which groups ought to be protected by hate speech laws. For more on these thorny issues, see Alexander Brown, *What is Hate Speech? Part 1: The Myth of Hate*, 36 L. & PHIL. 419 (2017); Alexander Brown, *What is Hate Speech? Part 2: Family Resemblances*, L. & PHIL. (online first, https://link.springer.com/article/10.1007/s10982-017-9300-x); Alexander Brown, *The "Who?" Question in the Hate Speech Debate: Part 1: Consistency, Practical, and Formal Approaches*, 29 CAN. J.L. & JUR. 275 (2016); Alexander Brown, *The "Who?" Question in the Hate Speech Debate: Part 2: Functional and Democratic Approaches*, 30 CAN. J.L. & JUR. 23 (2017).

such law, and although I can see something of that same care in Weinstein's article, such as when he distinguishes between different forms of hate speech law based on relative detriment to the legitimacy of so-called downstream laws, in some instances this care is lacking. Second, Weinstein plays up the importance of collective authorization or democratic legitimacy of downstream laws *vis-à-vis* "(a) the obligation of those restrained by the speech restriction to obey a downstream antidiscrimination law; and (b) the morality of enforcing the downstream measure against those whose participatory rights have been impaired by the upstream speech restriction."[4] These may be important aspects of what it means to detract from the legitimacy of downstream laws, but they do not exhaust the relevant aspects. Third, I think that Weinstein's article ignores some important nuances in what I have argued about hate speech laws and political legitimacy, and ignores something that might be true of the relationship between political and democratic legitimacy, namely, it might be that political legitimacy takes lexical priority over and, therefore, cannot be traded off against, the collective authorization or democratic legitimacy of downstream laws. Finally, I believe that in describing my use of the precautionary principle as "plainly indefensible" Weinstein has done justice neither to the raw plausibility of that principle nor to how I applied it to the special silencing effects of hate speech.

I plan to make the aforementioned points in the course of responding to two main objections that Weinstein levels against the arguments I made concerning hate speech regulations and political legitimacy in Chapter 7 of my book. The first objection concerns my response to Dworkin's argument that if we introduce "upstream" hate speech regulations and thereby "intervene too soon in the process through which collective opinion is formed," then "we spoil the only democratic justification we have for insisting that everyone obey these [downstream] laws."[5] I argued in response to Dworkin that there might also be a sense in which hate speech bans are not a threat to but a requirement of political legitimacy. I suggested that the question of the political legitimacy of, say, the legal system, might turn on whether the legal system could be the subject of interpersonal justification and consensus

4. Weinstein, *supra* note 1, at 535.

5. RONALD DWORKIN, *Foreword*, *in* EXTREME SPEECH AND DEMOCRACY v, viii (Ivan Hare & James Weinstein eds., 2009).

among free and equal citizens. More precisely, I said "that political legitimacy, including the legitimacy of the legal system, itself depends upon its being possible, at least in principle, to justify that system to each citizen bound by it on the basis of fundamentals of justice that they cannot reasonably reject."[6] I also proposed that

> members of minority or vulnerable groups could reasonably reject the following justification of an absolutist free speech doctrine. "For fear that hate speech law may put at risk the collective authorization . . . of downstream laws from which you benefit, we shall neglect to utilize the measures at our disposal to curb forms of hate speech that can be corrosive of a shared, public sense of the basic elements of your reputation, status and dignity as members of society in good standing."[7]

I believe that this attempt to justify an aggressive free speech regime to the victims of hate speech would fail because they would rightly see it as violating fundamentals of justice. By "fundamentals of justice" I mean, following Waldron's definition, "propositions establishing everyone's right to justice and elementary security, everyone's claim to have their welfare counted along with everyone else's welfare in the determination of social policy, and everyone's legal status as a rights-bearing member of society."[8] These are, I believe, basic propositions that everyone can, and should, be willing to accept, and that, under certain circumstances, will constitute grounds for reasonably rejecting an aggressive free speech regime.

However, Weinstein objects that despite my having provided a legitimacy-based argument for hate speech bans, I have nevertheless failed to provide a *like-for-like* legitimacy-based argument in response to Dworkin. In order for a legitimacy-based argument to have traction against Dworkin's contention that upstream hate speech bans can spoil the legitimacy of downstream laws then, such an argument must also work at the level of the legitimacy of downstream laws, claims Weinstein. In his words,

> even on the assumption that failure to enact hate speech laws does compromise legitimacy, it is, as Brown notes, the

6. BROWN, *supra* note 2, at 208.

7. *Id*.

8. Jeremy Waldron, *Dignity and Defamation: The Visibility of Hate*, 123 HARV. L. REV. 1596, 1626, n. 127 (2010).

> legitimacy of "the legal system" that has been diminished, not the obligation to obey or the morality of an enforcement of a particular law or laws. . . . It is difficult, however, to weigh a loss to systemic legitimacy against a detriment to the legitimacy of a particular law. The work done by these two types of legitimacy is very different. The concern of systemic legitimacy is, as Brown notes, identification with the legal system. In contrast, the concern about the legitimacy of a particular law that I have emphasized in this article is whether it is moral for the state to use force to make dissenters comply with a law with which they can reasonably disagree.[9]

Weinstein has missed some important nuances in what I said about political legitimacy, however. For one thing, what I actually said was that "political legitimacy, *including* the legitimacy of the legal system," itself depends on interpersonal justification and consensus among free and equal citizens. I used the term "including" in a non-exhaustive way, to mean *at least this (but not necessarily only this)*. Indeed, the illustrative example I gave focused on the interpersonal justification of what I called "an absolutist free speech doctrine," which is only one feature of the system of law, albeit an important one. It is an open question whether a failure to justify this feature would constitute not just a deficit in the legitimacy of this feature but also a deficit in the legitimacy of the entire legal and political system. This would depend on whether the legal and political system as a whole could be the subject of interpersonal justification and consensus among free and equal citizens, given attempts to justify the totality of its constitutional laws, civil rights laws, public policies, and so on. At any rate, I believe that the question of political legitimacy based on interpersonal justification and consensus among free and equal citizens can be meaningfully applied to particular features of the legal and political system, including both upstream and downstream laws, as well as to the entirety of that system.

This clarification is important for understanding what I would say about the legitimacy of downstream laws such as those involved in Waldron's English landlord example. Waldron asks us to image a landlord who discriminates against English families of South Asian descent in a way that is prohibited by English antidiscrimination laws. At the same time, English hate speech laws, such as laws banning the stirring up of racial hatred, prevent

9. Weinstein, *supra* note 1, at 577.

the landlord from using threatening, abusive or insulting words or behaviour with either the intention or likelihood of stirring up hatred against Pakistanis defined as a racial, ethnic or national group.[10] Whereas Dworkin claims that upstream laws can spoil the legitimacy of downstream antidiscrimination laws—laws that protect the very people who are also protected by the upstream laws—Waldron contends that "if we had a law that was specifically tailored to prohibit only expression at the viciously vituperative end of this spectrum, it might be an open question whether it would have anything more than a minimal effect on legitimacy."[11] Weinstein criticises Waldron for intimating that the detriment to the legitimacy of the downstream law could prove to be "minimal."[12] But he also criticises me for failing to provide a like-for-like legitimacy-based argument in response to Dworkin.[13] However, I believe that my account of political legitimacy *does* have the wherewithal to say something about the political legitimacy of downstream laws based on interpersonal justification and consensus among free and equal citizens. Specifically, I think that it is quite possible to justify an antidiscrimination law even to those people who disagree with it and who are denied certain specific types of opportunities (but not all types of opportunities) to publicly argue against it. It might go something along these lines. "You have an obligation to obey antidiscrimination laws, and we have a moral right to enforce antidiscrimination laws, for the simple reason that the state has a duty to fight injustice and it is clearly unjust to discriminate against people in their access to jobs, housing, transport, services, and so forth, merely because of their possession of protected characteristics, and, what is more, you have this obligation, and we have this moral right, even if hate speech laws reduce to some extent the collective authorization of these very same antidiscrimination laws and therefore diminish the democratic legitimacy of these laws, keeping in mind the fact that we are

10. Waldron, *supra* note 8, at 1643.

11. *Id.* at 1646.

12. Weinstein, *supra* note 1, at 532, and subpart A of Part III. Note, however, that whilst Weinstein claims these hate speech laws are detrimental to legitimacy, he stops short of saying that such laws would actually remove the landlord's normative obligation to obey the law and the state's moral right to enforce the law against him. Ibid. So despite Weinstein's baulking at Waldron's use of the word "minimal," both he and Waldron are in perfect agreement that whatever the nature of the detriment to the legitimacy of downstream law, it is not 'catastrophic.' Waldron, *supra* note 8, at 1642.

13. Weinstein, *supra* note 1, at 575–78.

utilising narrowly framed hate speech laws to curb forms of hate speech that can be corrosive of a shared, public sense of the basic elements of people's equal status and dignity as members of society in good standing, and corrosive of people's sense of physical security." I believe that such an attempt to interpersonally justify the antidiscrimination law would succeed because it appeals to fundamentals of justice that nobody can reasonably reject.

There is another, related nuance that Weinstein has missed. In my book I presented the process of interpersonal justification and consensus among free and equal people as a way of assessing hate speech law from "the sole perspective of political legitimacy."[14] I made it clear that "[t]his is not about trading off political legitimacy with the assurance of civic dignity but about the way in which the assurance of civic dignity is constitutive of the realization of political legitimacy."[15] In a similar vein, pace Weinstein's interpretation, I am *not* attempting to "weigh a loss to systemic legitimacy against a detriment to the legitimacy of a particular law." Rather, I am claiming that the upstream hate speech law is politically legitimate only insofar as it could be the subject of interpersonal justification and consensus among free and equal people and, similarly, that the downstream antidiscrimination law is politically legitimate only insofar as it could also be the subject of interpersonal justification and consensus among free and equal people. This holds true even if the downstream law suffers diminished *democratic* legitimacy due to the politically legitimate upstream law. So this is not about weighing a loss to systemic legitimacy against a detriment to the legitimacy of a particular law. Rather, it is about recognising the appropriateness of iterative applications of the test of political legitimacy for the legal system but also for upstream and downstream laws, and, furthermore, about placing *political* legitimacy and *democratic* legitimacy in what might be their rightful order of priority.

Weinstein's second objection focuses directly on my argument that free and equal people could reasonably reject a proposal for an aggressive free speech regime that disallows hate speech bans, even if the justification for the proposal appealed to

14. BROWN, *supra* note 2, at 208.
15. *Id*.

the protection of the democratic legitimacy of downstream laws. According to Weinstein, this argument underestimates the available evidence on the negative impact of hate speech laws on the democratic legitimacy of downstream laws. He writes:

> If this article has demonstrated anything, it is that hate speech laws as they actually exist, and of the type that Brown thinks justified, present much more than some "risk" to "collective authorization and legitimacy of downstream laws" from which members of these groups benefit. Nor, contrary to Brown's exposition of Waldron's erroneous view, have they resulted in only "relatively minor reduction in the collective authorization of downstream laws." Rather, as discussed in subsections B and C of this Part [IV], their effect on legitimacy, both in the normative and descriptive sense, is substantial. In light of such significant detriment to political legitimacy, even if one accepts hypothetical consent as the basis of political legitimacy, there is a very real question whether Brown's hypothetical interlocutors could reasonably consider the failure of a jurisdiction to enact broad hate speech prohibitions of the type Brown defends as contrary to "the fundamentals of justice."[16]

Now I am not entirely certain what Weinstein has in mind when he says I defend "broad hate speech prohibitions." But he does give a clue in the footnote attached to the sentence "hate speech laws as they actually exist, and of the type that Brown thinks justified." He writes:

> Thus far beyond the ban on highly vituperative hate speech that Waldron thinks might be justified, Brown defends bans on group defamation (*sensu stricto*) and on incitement to racial hatred, *id.* at 214. Despite the seemingly limited scope of such laws, they have, as I have demonstrated, been used to impair, and perhaps in some cases destroy, the legitimacy of downstream antidiscrimination laws. See subparts A and B of this Part [IV].[17]

Nevertheless, I find this to be a strange line of objection given that subparts A and B of Part IV of Weinstein's article in fact have nothing to say about group defamation laws (*sensu stricto*) and relatively little to say about incitement to hatred laws (aside from one Dutch case heard by the European Court of Human Rights (ECtHR) about which I shall say more in a moment). Instead, the vast majority of what he says in these subparts concerns

16. Weinstein, *supra* note 1, at 576.
17. *Id.* at 576 n.174.

expression-oriented hate crimes,[18] specifically, public order offences involving threatening or abusive words or behaviour, that are aggravated by hostility toward people based on their possession or perceived possession of certain protected characteristics.[19] In other words, despite what Weinstein suggests, subparts A and B of his article have certainly not "demonstrated" that group defamation laws (*sensu stricto*) and incitement to hatred laws have been used to impair or even destroy the legitimacy of downstream antidiscrimination laws.

To expand on this point, I take it that, for Weinstein, the most problematic hate speech laws are those that leave little or no leeway or room for people to express certain views in other permissible ways. In subpart A (Part IV), he cites various aggravated public order offences in England and Wales that he believes have, as applied, effectively prevented people from expressing in public their sincerely held religious view that homosexuality is immoral even without using epithets or slurs or stirring up hatred.[20] Then, in subpart B (Part IV), Weinstein spells out what he takes to be the "annihilation" of legitimacy of downstream laws associated with such upstream laws. Here he claims that people who have been prevented by such upstream laws from expressing in public the view that homosexuality is immoral no longer have a political obligation to obey downstream antidiscrimination laws. What is more, he claims that such hate speech laws might even, in worst case scenarios, "render immoral" the enforcement of downstream laws against people that have been silenced by them. However, these particular sorts of hate speech laws, what I call expression-oriented hate crime laws, are in fact much broader and more restrictive of speech than

18. See BROWN, *supra* note 2, at 35–38.

19. For some concrete examples of how such hate speech laws have been used in England and Wales, see Brown, *supra* note 3, at 285–86, 288–89 n.172, 311 n.252.

20. Two offences are involved in many of Weinstein's examples. The first is s. 4A of the Public Order Act 1986 ("Intentional harassment, alarm or distress"), which makes it an offence for someone to use threatening, abusive or insulting words or behaviour or to display any writing, sign or other visible representation which is threatening, abusive or insulting, with intent to cause harassment, alarm or distress and thereby causing harassment, alarm or distress. The second is s. 5 of the Public Order Act 1986 (as amended by s. 57 of the Crime and Courts Act 2013) ("Harassment, alarm or distress"), which makes it an offence for someone to use threatening or abusive (as amended) words or behaviour or to display any writing, sign or other visible representation which is threatening or abusive within the hearing or sight of a person likely to be caused harassment, alarm or distress thereby.

the group defamation laws (*sensu stricto*) and incitement to hatred laws that I defended in Chapter 7 of my book. So even if these expression-oriented hate crime laws have, in their application by the police, prosecutors and courts in England and Wales, effectively prevented people from even temperately expressing in public their view that homosexuality is immoral, the same is not necessarily true of incitement to hatred laws in England and Wales. Indeed, I would argue that these other, more narrowly framed hate speech laws *do* allow space for people to express certain views in other permissible ways.[21]

Thus, consider Part 3A of the Public Order Act 1986 (as amended by s. 74 of the Criminal Justice and Immigration Act in 2008), which *inter alia* sets out various offences relating to the stirring up of hatred on grounds of sexual orientation. It is narrowly framed in at least two important ways. First, it is written in such a way as to ensure that hate speakers have other permissible ways of stirring up hatred. This is because the law makes clear that the offences are only committed if people intentionally use "threatening" words or behaviour or written material or public performance of play or recording in order to stir up hatred. Non-threatening modes of expression, as well as unintentional stirring up, are untouched by these particular offences. Second, s. 29JA of Part 3A of the Public Order Act 1986 (as amended by Schedule 7 of the Marriage (Same Sex Couples) Act 2013) directly and explicitly states that people cannot be treated as stirring up hatred merely because they engage in discussion or criticism of homosexual conduct or gay marriage, for instance. This also gives speakers, including speakers motivated by sincerely held religious beliefs, some significant leeway to express their views on homosexuality and gay marriage without prosecution. I would add here that other countries also have similar caveats written into their incitement to hatred laws.[22]

21. I note that Waldron makes a similar point in his contribution to this symposium. *See* Jeremy Waldron, *The Conditions of Legitimacy: A Response to James Weinstein*, 32 CONST. COMMENT. 697, 700–04 (2017).

22. In Canada, for example, the part of the criminal code that bans wilful promotion of hatred also contains exemptions or permissible defences against prosecution "if, in good faith, the person expressed or attempted to establish by an argument an opinion on a religious subject or an opinion based on a belief in a religious text." *See* s. 319(3)(b) of the Criminal Code (as amended by An Act to Amend the Criminal Code (Hate Propaganda) of 2004). Likewise, some states in Australia have on the books legislation banning incitement to hatred which sets out exemptions for speech that has a religious purpose or is motivated by sincerely held religious beliefs. See ss. 11(b)(i) and 11(2) of the Racial and

Now, at this stage, Weinstein might point to how incitement to hatred laws have been used by the courts in some cases to *limit* or *reduce* the range of options for expression among hate speakers, and that this alone is a significant detriment to the democratic legitimacy of downstream laws, even if they have not been banned from expressing certain views as such. For example, in subpart A (Part IV) Weinstein does cite one example of an application of incitement to hatred law. In *Glimmerveen and Hagenbeek v. Netherlands* (1979)[23] the ECtHR judged as inadmissible applications made by two Dutch nationals who had been found guilty by domestic courts of possessing, with intent to distribute, leaflets that incited racial discrimination. So if what Johann Glimmerveen really wanted to do, as an exercise of his right to contribute to public discourse, was to express his views in such a way that constitutes incitement to racial discrimination, then he was not able to do so. In that sense his range of options were limited or reduced. Or consider the English case *R. v. Ali, Javed, and Ahmed* (2012),[24] which is not discussed by Weinstein. In July 2010, three devout but also socially conservative members of the Muslim faith distributed leaflets on the streets of Derby titled "Turn or Burn," "GAY – God Abhors You," "Death Penalty?," as a protest to the Gay Pride Festival taking place that day. They became the first people to be successfully prosecuted for offences relating to stirring up hatred on grounds of sexual orientation in England and Wales. In his sentencing remarks Judge Burgess made reference to the aforementioned clause 29JA, but nevertheless supported the jury's decision that in this particular case the wording of the leaflets *did* amount to the use of threatening words or behavior with the intention of stirring up hatred, based on the fact that four homosexual men had read the leaflet and "[a]ll felt threatened."[25] In that sense the intervention by the police and courts did limit or reduce these religionists' range of options in how they could permissibly express their view that homosexuality is a sin punishable in ways indicated in their religious texts. Maybe what Weinstein would say about this case is exactly what he says about some of the cases he presents in

Religious Tolerance Act 2001 (Vict.) and see ss. 80G(1)(b)(i) and 80G(1)(b)(i) (as amended by s. 6 of Law No. 80 of 2004) (W. Austl.).

23. App. No. 8348/78 & 8406/78, 18 Eur. Comm'n H.R. Dec. & Rep. 187 (1979).

24. No. T20110109 (Derby Cr. Ct.) (involving offences of stirring up hatred on grounds of sexual orientation).

25. Transcript obtained directly from Judge Burgess.

subpart A, Part IV, namely, that these speech restrictions "effectively prevented these citizens from participating in the public discussion of a host of antidiscrimination measures, as well as of proposals to extend marriage to include same-sex couples, in an intellectually honest and authentic manner."[26]

However, I believe that even if incitement to hatred laws do prevent some people from participating in public discourse in an intellectually honest and authentic manner, that is, in a manner of their choosing or in ways that perfectly express who they are as people and what they believe in, and even if this prevention thereby has a detrimental impact on the collective authorization and democratic legitimacy of downstream laws, free and equal people would nevertheless still have grounds to reasonably reject a failure to enact and apply such laws, which is a matter of *political* legitimacy. Whether the rejection is reasonable depends not simply on what they are rejecting but also on the grounds of, or reasons for, that rejection. To understand why free and equal people could reject even a free speech regime that aggressively protects the democratic legitimacy of downstream laws, it is necessary to comprehend the gravity of the relevant grounds or reasons. It seems to me that free and equal people might reasonably look upon the adequate protection of their equal civic dignity, such as via group defamation laws (*sensu stricto*), as a precondition of any notional agreement to joining the political community. This precondition is no less than what free and equal people would stipulate as the sort of basic status or standing they must retain in order for them to be willing to join together to form a political community with everything this joining together also entails about submitting to governmental institutions and a system of law that assumes an obligation to obey the laws because they are the laws and that claims a moral right to enforce the laws. Perhaps there are other fundamentals of justice, such as safeguarding people's sense of their physical security, that is, freedom from legitimate fear of acts of discrimination or violence, that are also preconditions for any notional agreement to joining the political community, and that would also require laws, including incitement to hatred laws, that combat hate speech that contributes to a climate of fear.[27]

26. Weinstein, *supra* note 1, at 562.
27. BROWN, *supra* note 2, at 66–75.

Now suppose for the sake of argument that I am correct to say that a regime of free speech that disallowed group defamation laws (*sensu stricto*) and incitement to hatred laws could be reasonably rejected on the grounds that it permitted the sorts of hate speech that can jeopardise people's sense of their equal civic dignity and even their sense of physical security. What does this mean for the political legitimacy of a political community that routinely strikes down such hate speech laws? In brute terms, it means that the community is less politically legitimate than it could be. But what implications follow from this *vis-à-vis* characteristic aspects of political legitimacy? Following Weinstein's lead, we might consider (a) a lesser normative obligation to obey certain laws and (b) a lesser moral right of the state to enforce certain laws. These do not, however, exhaust the possibilities. Take also (c) people having grounds on which to regret the loss of legitimacy and to strongly condemn the government concerned, (d) a right to engage in acts of civil disobedience short of disobeying justifiable laws, (e) a lesser obligation to support the system of law as a whole, once again short of breaking justifiable laws (for example, tax avoidance or arranging one's financial affairs to minimise tax liability within the law), and even (f) a lesser obligation to refrain from taking the law into one's own hands in the sense of enforcing hate speech norms that ought to be enshrined in law but are not (for example, making threats of extrajudicial punishment against hate speakers). Unfortunately, I do not have up my sleeve a theory that can easily tell us which of these implications are most fitting for political communities that unjustifiably fail to enact and apply incitement to hatred laws, for example. But I do want to make the point that no adequate discussion could begin and end with an assessment of (a) and (b) alone.

I now turn to Weinstein's second main objection. This objection concerns the fact that in my book I also invoked the precautionary principle in order to justify, or supply further justification for, certain forms of hate speech law. Specifically, I argued that

> an authority may adopt laws forbidding hate speech when it amounts to discriminatory harassment in the workplace or on campus, or laws interdicting hate speech when it constitutes discriminatory intimidation, because having identified the possibility of the catastrophic antidemocratic outcome that a proportion of the individuals targeted by hate speech will not

> participate in the formation of public opinion, and bearing in mind the conditions of uncertainty that surround these outcomes, it errs on the side of precaution.[28]

I was certainly not the first scholar to appeal to the precautionary principle as a justification for the regulation of hate speech. In her 2008 article, "A Constitutional 'Right' to Deny and Promote Genocide?," for example, Karen Eltis appealed to the principle as a way of both reinterpreting and defending the Supreme Court of Canada's decision in *R v. Keegstra*.[29] This case involved a prosecution of an openly anti-Semitic school teacher for the crime of wilful promotion of hatred under s. 319(2) (ex s. 281.2(2)) of the Criminal Code.[30] According to Eltis, "the majority opined that hate speech can serve as a precursor to genocide and, through its ruling, advocated a precautionary approach to denial and incitement."[31] A similar precautionary justification also seems to have played a part in the thinking of the Kenya National Commission on Human Rights (KNCHR) in the course of its work monitoring the 2005 Constitutional Referendum and 2007 General Elections and in pushing through new incitement to hatred laws in Kenya.[32] It is also implicit in Mari Matsuda's most recent article on hate speech.[33] Genocide is not the only applicable harm. For instance, Richard Posner has argued that incitement to racial and religious hatred might constitute a long term threat to national security of sufficient magnitude to warrant legal sanctions even in the U.S.—on the evidence that heightened levels of incitement to hatred against American Muslims increases the risk over the long term of terrorist attacks by American Muslims on home soil—even if we do not know exactly how long it would take for the risk of a terrorist attack to significantly ramp up.[34] Whilst Posner does not endorse the precautionary principle,

28. *Id*. at 199.

29. [1990] 3 S.C.R. 697 (Can.).

30. Karen Eltis, *A Constitutional "Right" to Deny and Promote Genocide? Pre-empting the Usurpation of Human Rights Discourse Towards Incitement from a Canadian Perspective*, 9 CARDOZO J. CONFLICT RESOL. 463 (2008).

31. *Id*. at 476.

32. *See* Lawrence Murugu Mute, *Legislation, Hate Speech, and Freedom of Expression in Kenya*, PAMBAZUKA NEWS (Oct. 22, 2008), www.pambazuka.org/governance/legislation-hate-speech-and-freedom-expression-kenya.

33. Mari J. Matsuda, *Is Peacemaking Unpatriotic?: The Function of Homophobia in the Discursive World*, 11 J. HATE STUD. 9, 9–10 (2013).

34. RICHARD POSNER, NOT A SUICIDE PACT: THE CONSTITUTION IN A TIME OF NATIONAL EMERGENCY 124 (2006).

what he is suggesting is a sort of cost-benefit analysis that builds in a margin of safety when it comes to risks of temporally distant but especially serious harms.[35]

The sorts of harms that Eltis, the KNCHR, and Posner have in mind (genocide, terrorist atrocities) are grave and irreversible. They are equivalent to the devastating climate change harms that are associated with the precautionary principle in the field of environmental regulation. Of course, the sorts of harms that I focused on in Chapter 7 of my book are not of the same magnitude of gravity as these. Nevertheless, they are potentially more probable harms and more proximate harms, causally speaking. While they are not strictly irreversible (as with loss of life), they are not easily reversible. And whilst less grave, they are still extremely serious. I am speaking of the antidemocratic outcome that a proportion of the individuals targeted by hate speech will not contribute to public discourse nor participate in the formation of public opinion, or will do so but with speech the content of which has been warped, or will do so but without the ability or power to achieve intended illocutionary or perlocutionary effects.[36] (Of course, the precautionary principle might also be applied to other types of hate speech law that address other categories of extremely serious harm other than the aforementioned antidemocratic outcome.[37]) According to Weinstein, however, my "invocation of the precautionary principle in lieu of evidence . . . turns a problematic though plausible argument into a plainly indefensible one."[38]

35. *Id*. at 122. *See also* RICHARD POSNER, CATASTROPHE: RISK AND RESPONSE (2004).

36. For an overview of the literature, see BROWN, *supra* note 2, at 84–86, 198.

37. For example, the principle could potentially be invoked to justify laws disallowing the use of epithets or insults directed at or targeted against individuals based on their possession or perceived possession of protected characteristics, such as if such speech had the potential to cause in a proportion of those subject to it serious psychological damage, such as anxiety or distress, or to exacerbate the symptoms of pre-existing mental illnesses, such as depression, or to trigger the onset of mental illnesses to which individuals may have already been at risk, such as antisocial personality disorder, even if there was a lack of decisive or consensus-based evidence to prove these effects. *Id*. at 49–58. Or, to take another example, the principle might be invoked to justify laws banning incitement to hatred if such speech had the potential to significantly contribute to the production and maintenance of a climate of hatred which is partly constituted by an increased chance of acts of discrimination, violence, damage to property, and so forth, even if there was a lack of decisive or consensus-based evidence to prove this contribution. *Id*. at 66–75.

38. Weinstein, *supra* note 1, at 580.

According to the strong version of the precautionary principle I had in mind, the main burden of proof is placed not on those wishing to regulate or support the regulation of potentially harmful activities but on those who wish to engage in or support those activities. It is they who should demonstrate that the activities would not produce significant harms if left unregulated, based on evidence that is sufficiently rigorous, comprehensive and abundant to command a consensus among the relevant body of experts. The upshot is that the principle may require regulation of activities because there is lack of consensus-based evidence that the activities would not produce significant harms if left unregulated. It goes without saying, however, that there must be at least some minimally adequate evidence that the relevant activities have certain effects and that these effects are potentially harmful in order to shift the burden in this way. I believe that this threshold has been met for hate speech and various types of silencing effect.[39] But it is not necessary for the evidence of harm itself to be consensus-based.

At any rate, the crux of Weinstein's objection to my invocation of the precautionary principle is his assertion that there is a worrying asymmetry between the regulation of hate speech that has *certain* or *known* harmful effects, measured in terms of some speakers' reduced opportunities to contribute to public discourse and participate in the formation of public opinion, and the justification of such regulation in the name of preventing only *possible* or *potential* harmful effects, measured in the same way (only focusing on the subjects of hate speech). He writes:

> The view that bigots can be forbidden by force of law from expressing their views—which will, if the law has any effect at all, undoubtedly have a "silencing effect" on them—to avoid *the possibility* that some unspecified "proportion of the

39. In other words, I do not accept the premise (present in some objections to hate speech laws) that there is a paucity of evidence (i.e., not a minimally adequate level of evidence) of harmful silencing effects. *See* BROWN, *supra* note 2, at 98–99. For evidence of silence (or passivity) as a common response to, and effect of, hate speech, see, e.g., Laura Leets, *Experiencing Hate Speech: Perceptions and Responses to Anti-Semitism and Antigay Speech*, 58 J. SOC. ISSUES 341 (2002), and Katharine Gelber & Luke J. McNamara, *Evidencing the Harms of Hate Speech*, 22 (3) SOC. IDENTITIES 324 (2016). For an analysis of what silencing means in terms of harmfully removing real opportunities to participate in the formation of public opinion, see, e.g., BROWN, *supra* note 2, at 194–208. I would like to thank the journal's editor, Jill Hasday, for suggesting I make these important clarifications, which I had not made in the original version of my contribution.

> individuals targeted by the hate speech" *might be* deterred from speaking is simply impossible to square with the basic premise underlying participatory democracy that all citizens should have the equal opportunity to engage in the formation of public opinion regardless of the viewpoint they want to express.[40]

Of course, the known harmful effects of hate speech regulations are not limited to reducing the speakers' opportunities to contribute to public discourse and participate in the formation of democratic public opinion.[41] But, to focus on the democratic harms, I believe that in addition to the asymmetry between certain and potential effects, there is another asymmetry that Weinstein either ignores or underestimates, and that makes the invocation of the precautionary principle more not less justified. The asymmetry I have in mind is in the nature of the silencing effect at issue. When the state intervenes to criminalise certain forms of hate speech, narrowly framed laws will curtail only that given form of speech. Such laws, sensibly and properly applied, will not stop the speaker from expressing him or herself in other permissible ways. As mentioned above, if laws prohibit the use of threatening words or behaviour to stir up hatred, then hate speakers can perform the same speech acts in other ways, using other kinds of words or behaviour. So, for example, a hate speaker might be banned from saying this. "You think you can trust Muslims, think again, they are vile, backward, and dangerous people who deserve only our hatred, and when this country is finally united in its hatred of Muslims, they had better watch out!" But he might not be banned from saying this. "You think you can trust Muslims, think again, they are vile, backward, and dangerous people who deserve only our hatred." By contrast, in the event that hate speech has a silencing effect on those who are its subjects, the effect is just that, silence; it can cause people not to speak in any way. Of course, Robert Post is very convincing when

40. Weinstein, *supra* note 1, at 580.

41. Other harmful effects include a reduction of negative freedom to engage in hate speech and thereby to pursue self-development (for example, truth discovery, self-realisation). *See* BROWN, *supra* note 2, at ch. 4. Then there is the loss of formal autonomy. Hate speech regulations substitute a governmental choice for a personal choice about how and when to embody one's values in speech, including a personal choice as to what is appropriate and what is inappropriate public speech. *See*, *e.g.*, C. Edwin Baker, *Hate Speech*, *in* THE CONTENT AND CONTEXT OF HATE SPEECH: RETHINKING REGULATION AND RESPONSES 57, 63–64 (Michael Herz & Peter Molnar eds., 2012). For further discussion, see BROWN, *supra* note 2, at 210–13.

he states that part of the point of the First Amendment is to protect citizens' *right to choose* the ways, manner, and circumstances of their participation in the formation of public opinion.[42] But surely another, perhaps even more fundamental purpose of a regime of free speech is to ensure that *all* citizens enjoy at least sufficient real opportunities to participate in public discourse.[43]

It seems to me that part of the raw intuitive appeal of the precautionary principle stems from the idea that we ought to be better safe than sorry. When it comes to hate speech harms it is also important to attend closely to who the "we" are, and what harm they face. It is precisely because the nature of the harm (silencing) is different for hate speakers and those people who are the unwilling subjects of hate speech that it is *not* impossible to square the precautionary approach with the principle that all citizens should have sufficient real opportunities to engage in the formation of public opinion. Perhaps Weinstein thinks that any application of a strong version of the precautionary principle is plainly indefensible. If so, then the grounds for his objection takes him well beyond the hate speech debate and, more importantly, his objection requires significant bolstering, well beyond his brief remarks in the article.[44]

Building on his two main objections, Weinstein ends with an offer to rescue my arguments. He writes: "I will build on some of Brown's better arguments to try to identify a countervailing legitimacy concern sufficiently similar in type and character to the

42. *See*, *e.g.*, Robert Post, *Reconciling Theory and Doctrine in First Amendment Jurisprudence*, *in* ETERNALLY VIGILANT: FREE SPEECH IN THE MODERN ERA 153, 167 (Lee C. Bollinger & Geoffrey R. Stone eds., 2002); Robert Post, *Democracy and Equality*, 1 L., CULTURE & HUMAN. 142, 148 (2005).

43. *See* BROWN, *supra* note 2, at 194–201.

44. For general criticisms of the precautionary principle in various spheres of law and regulation, see CASS SUNSTEIN, LAWS OF FEAR: BEYOND THE PRECAUTIONARY PRINCIPLE (2005); Stephen G. Wood et al., *Whither the Precautionary Principle? An American Assessment from an Administrative Law Perspective*, 54 AM. J. COMP. L. 581 (2006); and STEVE FULLER & VERONIKA LIPINSKA, THE PROACTIONARY IMPERATIVE: A FOUNDATION FOR TRANSHUMANISM (2014). For a defence, see Timothy O'Riordan & Andrew Jordan, *The Precautionary Principle in Contemporary Environmental Politics*, 4 ENV'T. VALUES 191 (1995); Marko Ahteensuu, *Defending the Precautionary Principle Against Three Criticisms*, 11 TRAMES 366 (2007); David A. Dana, *The Contextual Rationality of the Precautionary Principle*, 35 QUEEN'S L.J. 67 (2009); and Nassim Nicholas Taleb et al., *The Precautionary Principle (With Application to the Genetic Modification of Organisms)* (NYU Sch. of Engineering Working Paper Series, Sep. 4, 2014), https://arxiv.org/pdf/1410.5787.pdf.

legitimacy that I have argued is diminished or destroyed with respect to downstream legislation, and which, therefore could, at least theoretically, offset this deficit."[45] He then proffers the following interesting case.

> Suppose, for instance, that in a certain democratic country the legislature is considering whether to grant an exemption from its drug laws to members of an indigenous population to use a substance traditionally employed by this group in religious ceremonies. Suppose that it is also the case that hate speech against this group, long subject to discrimination by the European settlors and their descendants, is so rampant and virulent that many members of this vulnerable minority group are "out of fear for their personal safety or livelihood" reasonably deterred by the hate speech from publicly supporting the exemption. If the exemption is not passed, then members of this indigenous community might well feel, and aptly so, that they have no political obligation to obey a law against ingesting the drug as part of their religious ceremony.[46]

However, Weinstein does not go on to actually defend what I attempted to defend, namely, banning the hate speech in question. He claims that "[i]t must be further demonstrated that the gain in legitimacy produced by the hate speech ban at least marginally exceeds the detriment to legitimacy caused by the speech restriction."[47]

I believe that Weinstein misunderstands what is really at stake here. For one thing, his talk of gains in democratic legitimacy produced by the hate speech ban offsetting the detriment to democratic legitimacy caused by the speech restriction is anathema to what I see as the basic proposition underlying participatory democracy and democratic legitimacy. The real touchstone is ensuring that *all* citizens enjoy at least sufficient real opportunities to contribute to public discourse and participate in the formation of public opinion. I also think that it makes perfect sense to ask whether or not this basic proposition or touchstone can itself claim *political* legitimacy, based on the test of interpersonal justification and consensus among free and equal citizens. The question is: would free and equal people have reasons based on the fundamentals of justice to reject an aggressive free speech regime that treated hate speech as a

45. Weinstein, *supra* note 1, at 580.
46. *Id.* at 580–81.
47. *Id*. at 581.

protected category even though certain forms of hate speech carry a risk of effectively removing from some people who are the subject of hate speech real opportunities to contribute to public discourse and participate in the formation of public opinion?

Moreover, I believe that by refocusing the debate about hate speech law onto the question of obligations to obey downstream laws, Weinstein has overlooked a far more important question. Why should people who are subject to hate speech and interpersonal silencing be put in a position of having to think about whether they have no political obligation to obey downstream laws, much less of having to contemplate disobeying these laws as a means of addressing the diminished democratic legitimacy of those laws? To fall into this way of thinking about their predicament risks imposing three harms on the victims of hate speech: first, the harm of being subject to silencing hate speech, as in, hate speech that, due to its psychological as well as material effects, freezes them out of contributing to public discourse and participating in the formation of public opinion; second, having a decision about downstream laws go against them partly because they did not contribute to public discourse and participate in the formation of public opinion; and third, the harm of having to engage in potentially risky forms of civil disobedience simply to make their point. Adopting the aforementioned ideal of political legitimacy as well as the precautionary principle as justifications for effective and narrowly drawn hate speech regulations means that victims of hate speech may be spared this triply unenviable position.