State Speech vs. Hate Speech:
What to Do About Words that Wound?

Abstract

This is, indeed, another work on the subject of hate speech regulation in the United States. And yet, it is not just another such work. For my goal here is not to settle the jurisprudential arguments regarding the possibility of any specific hate speech regulation, either extant or yet to be conceived, withstanding a Constitutional test. Nor is it my intention to demonstrate, on the basis of a comparative study of existing legislation, that such regulation either is or is not effective in addressing or redressing the social ills of hatred, discrimination, and inequality. Rather, I will achieve greater analytical clarity about just what the harms of hate speech are. I do so in order to reinvigorate the question about regulation with a new view of what exactly the object needing attention is, by demonstrating that though there are real harms here, the state cannot provide a regulatory remedy (at least qua criminal justice). Thus, in my conclusion I will assert that the question of what we might do differently in response to hate speech can only be answered—however provisionally—insofar as we first confront how we need to think differently about it. Specifically, I will argue that we need to replace the emphasis on redressing harms once they have occurred with a new emphasis on addressing, and ultimately eliminating, the conditions which make those harms possible in the first place.

Introduction

This is, indeed, another work on the subject of hate speech regulation in the United States. And yet, it is not just another such work. For my goal here is not to settle the jurisprudential arguments regarding the possibility of any specific hate speech regulation, either extant or yet to be conceived, withstanding a Constitutional test. Nor is it my intention to demonstrate, on the basis of a comparative study of existing legislation, that such regulation either is or is not effective in addressing or redressing the social ills of hatred, discrimination, and inequality. Rather, I will achieve greater analytical clarity about just what the harms of hate speech are. I do so in order to reinvigorate the question about regulation with a new view of what exactly the object needing attention is, by demonstrating that though there are real harms here, the state cannot provide a regulatory remedy (at least qua criminal justice). Thus, in my conclusion I will assert that the question of what we might do differently in response to hate speech can only be answered—however provisionally—insofar as we first confront how we need to think differently about it. Specifically, I will argue that we need to replace the emphasis on redressing harms once they have occurred with a new emphasis on addressing, and ultimately eliminating, the conditions which make those harms possible in the first place.

In order to meet my goals I will first, in §1, survey existing claims about the harm of hate speech;
both the views expressed by those who support regulation, and by those who do not. What we will find is that, despite more than a generation of legislation and court action on this issue, not to mention a great deal of theoretical attention to it over the past decade, we still not have managed to agree just what is the phenomenon “hate speech,” and what are its consequences, which we ought or ought not to regulate.

In the main body of the paper—in §§2 and 3, successively—I will address the three kinds of harm I believe all extant claimed harms can be classed under. In §2, I will discuss the intrinsic harms of hate speech; that is, the psychological and ontological deficits and difficulties that are embodied in the speech itself and/or in the ideas it carries. In §3, I will examine the implicit harms of hate speech; that is, the future physical violence that the speech either promises (in the form of threat, direct or oblique) or the former physical violence the speech endorses having occurred. In both sections, I will (a) analyze both (i) the theoretical status of the kind of the harm (that is, its logical validity and its ontological significance as an idea), as well as (ii) its empirical status (that is, its existence and significance as a political phenomenon), and (b) consider its regulatory costs (that is, the political, economic, and practical disadvantages to addressing this harm through different regulatory practices).

In the conclusion (§4), we will consider, in light of the evidence provided by the three-part analysis of each of the three harms identified, whether hate speech regulation is (a) possible, but expensive, or (b) simply impossible. That is to say, is the regulation of the harms inflicted on individuals and groups by hate speech merely costly (even very costly), like stipulating that the state provide 1800 calories of low-fat, high-fiber food to all people living within its territorial sovereignty each day? Or, is it just impossible, like said state stipulating that, within its territorial sovereignty, human beings only need 600 calories each day to subsist healthfully and productively? We will find that hate speech is, unfortunately, impossible to regulate, insofar as intrinsic harms are impossible to regulate and implicit harms are impossible to regulate any more or any better than they are through “fighting words” provisions and harassment laws.

But since it is entirely unsatisfactory, both on the theoretical level as well as on the practical level, to say that there is a distinct harm here, and that there is no possibility of a successful regulatory response—at least in the register of criminal justice, I will then, in the very final moments of the paper ask whether or not there is some other response by which we can address the harms I assert are perpetrated in hate speech without pushing for a criminal justice redress which isn’t viable. If I am right in laying out the unique harms of hate speech in the way I delineate them in §§2 and 3, then might we find within my recasting of the problem of hate speech a recasting of the possibility of its solution? We will see that we do in fact find a new approach in seeking a better resolution, which I call “striving to address, rather than failing to redress, the harms of hate speech.” I understand by this phrase a whole series of initiatives, both community-centered and state-supported, which would shift the emphasis from responding to hate speech when it happens to undermining the conditions for the possibility of the speech in the first place. This series of initiatives will be spelled out at the very close of the paper.

1. Is there really any harm in hateful speech?

The debate about whether or not the United States should join what many have seen to be a growing global consensus on the need for such regulations has been raging for more than a decade.
now, and yet the situation here seems very much the same today as when the debate started. There is still no federal legislation on the matter; though there are a number of “bias crime” laws on the books—more than 80 percent of the states have some such regulation. But those who believe there is something special and inimical about hateful speech, something which escapes prosecution so long as there isn’t legislation specifically criminalizing speech that carries a hateful content, are unsatisfied.

Mari J. Matsuda has laid out a definitive argument for criminalizing the most virulent forms of racist hate speech, an argument to which I am entirely sympathetic and with which I nearly entirely disagree. This account and the responses to it provide an excellent summary of the debate surrounding hate speech regulation today. Though her argument warrants more attention than it receives here, we can say most schematically that Matsuda wishes to prove three things in buttressing her pro-regulation conclusion. (1) Racist hate speech is violence, and specifically violence visited upon the most historically undermined persons in democratic societies. (2) Racist hate speech, therefore, as the world community of nations recognizes, is a crime that can (on reasonable grounds) and must (on normative grounds) be punished. (3) Despite appearances, statement (2) does not necessarily conflict with the First Amendment of the United States constitution.

Let’s address these claims one by one, beginning with Matsuda’s first central claim; namely: (1) Racist hate speech is violence, and specifically violence visited upon the most historically undermined persons in democratic societies. Judith Butler has argued, partially in response to Matsuda, what we might mean when “we claim to have been injured by language,” beginning with an analysis of name-calling, which she takes to be “one of the first forms of linguistic injury that one learns” (Butler, Excitable Speech, p. 2). What exactly is it that happens when one is called a name? For one thing, such a process is not always injurious. It is, indeed, one of the main modes in which we achieve subjectivity in an interpersonal context. This “mode” is what she calls, following Althusser, “interpellation.” Interpellation is of great importance for her analysis because she is pursuing the possibility that “the power of language to injure follow[s] from its interpellative power” (Butler, p. 2). What Butler is suggesting--however confounding and complicated it may sound--is really just that it is not the words themselves, but the mode in which they are expressed, that makes them powerful, injuriours, harmful. As she states: “[L]inguistic injury appears to be the effect not only of the words by which one is addressed but the mode of address itself, a mode--a disposition or conventional bearing--that interpellates and constitutes a subject” (Butler, p. 2). There is, that is, the words themselves cannot be called the sine qua non of the harm. Rather, the harm that is inflicted is derived from the mode, the manner, in which the words are spoken.

This becomes clearer in Butler’s direct response to Matsuda and other authors in her edited volume. She states: “To claim that language injures or, to cite the phrase used by Richard Delgado and Mari Matsuda, that ‘words wound’ is to combine linguistic and physical vocabularies. The use of a term such as ‘wound’ suggests that language can act in ways that parallel the infliction of physical pain and injury” (Butler, p. 4). Here we can already see the emerging problem with Matsuda’s account. She (like others) fails to distinguish between the kind of injury inflicted in language and the kind inflicted on the body. The point can be made clearer still by looking at a simile--and it is of special importance to Butler that it is a simile--drawn by Charles R. Lawrence III in the same volume, who states that “the effect of racial invective is ‘like receiving a slap in the face,’” as “the injury is
instantaneous’’ (Butler, p. 4). What is important here for Butler’s analysis is not the “slap in the face” or the “injury” which is “instantaneous.” Rather, she is focused on this “like.” After all, similes—comparisons that use “like” or “as”—compare unlike things. So there is an “unlikeness” between physical and linguistic injuries which these thinkers have not been able to isolate. This failure has established the further failure that they are unable to make a compelling argument for the “real harm” of “hate speech.” For, in failing to adequately expose the unlikeness between the harms of speech and the harms of a physical assault, these authors have failed to point out to us what is specific to (and specifically insidious about) hate speech as a medium of harm.

For all the cogency of this criticism of their use of simile, it is true that Butler isn’t really giving Matsuda and her fellow authors due credit here. Indeed, both Matsuda and Lawrence demonstrate that there are unique psychological and even physical dimensions to the experience of being subjected to hate speech. All this being what it is, the point for us remains that the interpellation of a less powerful (indeed, often powerless) human being by another in a position of power can only be understood by appreciating what Butler is calling (following J.L. Austin) the “total speech situation” (Butler, p. 3). It is in this context that an injury is quite certainly possible, and given the right circumstances—such as hatred or some other motivation for domination—likely. How this likelihood becomes a reality is the process by which the negative qualities attributed to the object of the hate speech (that is, the living, breathing human being who is being dehumanized) come to be believed by that object (person), so as to create an identification between the perpetrator and the victim.

The problem with this account is that, as the grammar of my last sentence shows, it presupposes precisely what needs to be proven: that there is an injury inflicted on some victim equal to the violence done (or attempted) by the perpetrator. That is, basing the argument for regulation on harm forces us to assume that the hate-mongerer has succeeded in harming the object of their hate. That their interpellative act was successful. But law cannot make such a presumption. In order to be convicted—even to be indicted—of (say) assaulting me, it has to be clear that I have been assaulted; not merely that you might have wished or intended or attempted to assault me. So with hate speech: if it is criminal, it must be clear that the intended victim suffered a harm, not merely that the perpetrator wished for them to do so. Can such a harm be established? We will pursue this possibility further when we discuss the “intrinsic” harms of hate speech in §2.

One place to look in trying to get a more precise hold on this harm is in extant hate speech regulations. Matsuda recognizes this in her second claim: Racist hate speech, as the world community of nations recognizes, is a crime that can (on reasonable grounds) and must (on normative grounds) be punished. The clearest formulation of this argument comes when she states: “If the harm of racist hate messages is significant, and the truth value marginal, the doctrinal space for regulation of such speech becomes a possibility. An emerging international standard seizes this possibility” (Matsuda, ‘Public Response to Racist Speech’, p. 26). This statement displays the tight bind between her first and second claims: Matsuda is saying that since there is a real harm involved in this phenomenon of racist hate speech, and since there is little to no truth value to be found in it, what we have could be called a crime. And, she takes as evidence for this, the fact that the world is coming to a consensus that we are, in fact, talking about a crime when we talk about racist hate speech. This brings her to a survey of this sort of legislation around the world, as supported by U.N. task forces, conventions, and discussions, in order to show that such measures are ever more
popular, and effective.\textsuperscript{8}

On the face of it, Matsuda’s claim of a developing consensus for criminalizing hate speech seems only more true now than when she wrote a decade ago. In September 2003, in fact, Canada expanded its anti-hate legislation to protect persons from hate on the basis of “sexual orientation.” South Africa has also recently begun enforcing similar legislation.\textsuperscript{9} And Australia has recently developed its first commonwealth-wide regulation, based on an extant regulation used in New South Wales.\textsuperscript{10} But, on closer inspection, it turns out that none of this legislation gets us any closer to a clear definition of the harmfulness of hate speech, because they provide conflicting definitions of that harmfulness, or no account at all. And therefore, both extant legislation worldwide and anything adopted in the United States does and will suffer from a lack of clear application. This underscores more strongly the need to clarify the real nature of this harmfulness, the source of any putative criminality.

A further difficulty for Matsuda and all those who would advocate criminalization is that her third claim—that, despite appearances, statement (2) does not necessarily conflict with the First Amendment of the United States constitution—cannot be justified.\textsuperscript{11} In fact, there \textit{is} a conflict, and a deep one, between any hate speech regulation implemented in the United States and the First Amendment. The problem is this: at the very heart of the definition of speech in founding of the United States is the distinction between political speech and any other speech. So while all sorts of speech (like advertising, gossip, pornography, or “shouting fire in a crowded theater”) have found themselves liable to civil and criminal litigation over this nation’s history, political speech remains ensnared and seemingly unassailable within the Bill of Rights. Why does the proponent of hate speech regulation in the United States have to confront this distinction between political and other speech? Because it is on precisely political grounds that such regulation would be undertaken in the first place. It might be possible, that is, to frame a hate speech regulation that does not conflict with the protection of political speech—a regulation that protected individuals on the basis of their character or something peculiar to them (their dress or their breath, say). But what hate speech regulation wants to regulate is speech directed at individuals \textit{as members of particular (historically undermined) groups}.

But speech about individuals as members of such groups, when exercised by individuals expressing their independent views, is precisely political speech. It is one thing to speak about the state apparatus making claims about members of groups—this is propaganda and has been successfully prosecuted (although only after major, horrific devastation) both in Nuremberg and more recently in the case of the Rwandan genocide—but for individuals to express their views freely and openly about various constituencies has been considered an integral part of the political at least in the nation, and at least since its founding. It is of course possible that such speech is also slander or incitement to violence. In such cases, it should be prosecuted as such. But to make an \textit{a priori} determination that any such speech is prohibited is to foreclose the political contestation of and about individuals as members of groups. And this is precisely, among other things, what political deliberation is meant to open up, and certainly not foreclose in advance. Therefore the phenomena that regulation advocates have most strongly in mind are precisely the kind of phenomena (again, among others, to be sure) that the First Amendment is meant to protect. And so, the idea that a form of speech can be regulated because it has low truth content—which has been at the heart of the successful regulation of the kinds of speech noted above—simply does not have purview here; as it
is precisely intended that the citizenry must come to its own understanding of the true and the false when it comes to political debate.

Matsuda, of course, is not unaware of this. Her response—and that of advocates for regulation in the American context more generally—would be that the (arguably limited) violation of First Amendment protections necessitated by regulation must be countenanced because failing to do so means a much greater and much more devasting (and unjust) violation of the 14th Amendment. Legally considered, the assertion is that the equality principle advanced in the later Amendment trumps the free speech principle advanced in the former. Though it is beyond my scope to assess this assertion here, it is clearly a stretch at best on jurisprudential grounds.\textsuperscript{12}

Taken broadly, the assertion is that the norm of equality ought to be esteemed more highly, even if at the expense of the norm of free speech. More perspicuously, it is the claim that the positive prescription to achieve greater equality needs be given greater heed, \textit{vis-à-vis} the negative proscription of government’s ability to interfere with the speech acts (and especially the political speech acts) of individuals. This broader claim will be examined in light of the each of the harms identified in §§2 and 3, and then definitively in §4, where I will make two final arguments: (1) even though this broader claim has great credence, it does not justify regulation as criminalization; (2) nevertheless, it does point to some avenues those sympathetic to the claims of regulation advocates still have open to them in seeking to address the harms of hate speech.

2. The harms of hate speech, 1: “Intrinsic” harms.

A.

In trying to define the harm of hate speech, Martha Minow tells us,\textsuperscript{13} we must look at the hateful speech act from the perspective of members of groups traditionally subjected to this sort of offense. What precisely we see when we do so is what we called above “interpellation”: the moment in which your being is named for you by another. In Minow’s account this amounts to the “dehumanization” of the victim of hate speech. She reminds us that it is not merely the one who hates who holds a dehumanizing view of the object of hatred--\textit{this} person who ceases to be a person (who is “dehumanized”). Indeed, “the wrongdoer’s reference group also holds the dehumanizing view, and often, it circulates in the broader society as well” (Minow, p. 33). Shifting perspective allows us to see that the dehumanizing claim is inscribed not just in the hater--the perpetrator, the actor--but also in the hated--the victim, the patient. And once we see this, persons committed to deepening our collective commitment to the wrongness (in both of this term’s senses--i.e., ethical and epistemological) of this claim and its interpellation of (real, living) persons, cannot not act.

But, as I said above, this account of interpellation does not really grasp the psychological and physical consequences of being subjected to hate speech as a member of a historically disadvantaged group. After all, it is often the case that those who have been subjected to these speech acts say that they precisely were not interpellated. That they did not find themselves understanding themselves through the words—specifically, the hate-inspired dehumanizing claims—of their assailant. Given this, how might we better capture this harm, which I am here calling the \textit{intrinsic} harm of hate speech?

If such a harm exists, it must be located in the speech act itself and/or the existential conditions in
which the act occurs, and not in the psyche of the person subjected to it. Otherwise, we inevitably run into the objection I just raised regarding the fact that many “victims” of hate speech do not recognize such a harm. What then, is it about the speech and its context with which we can take issue (to put it mildly) given the fact that said speech is often unsuccessful in its dehumanizing (i.e., interpellative) intent?

Orville Lee believes that this harmfulness lies in the fundamental fact of the “force of words”—that is the “symbolic power” which serves “as the condition of the possibility for the social existence of individuals and groups” (Lee, p. 862). Linguisticity—or having an existence which is permeated by symbolic interaction, even at times when we are not hearing, saying, seeing, or thinking in language per se—is a basic fact of human existence. Therefore, the argument goes, no account of our social existence, including the “formation and contestation of social hierarchies” (Lee, p. 863), can reflect the possibilities (both negative and positive) of human interaction if it does not accurately describe this power, this force of words.

In this case, the crucial aspect of our linguisticity is our “linguistic vulnerability,” which Lee (following Butler) understands as a necessary precondition for our existence as social beings in a social context. There is no way to be recognizable within a given society without being subject to harm, or: “harm and recognition are two sides of being situated in language” (Lee, p. 865). But since this is an ontological condition, it cuts through social categories and existing hierarchies of advantage and disadvantage: “The subject’s vulnerability to the force of words…exists prior to the institutional configuration of speakers and linguistic meaning” (Lee, p. 864). While this does provide the “performativ space” for reappropriations of hateful words which Butler advocates, it also means that an argument resting on existing institutionalized inequalities in defining the harms of hate speech will miss the mark. For our broader purposes in diagnosing the prospects of regulating hate speech, this means that the “equality clause” arguments for regulation are ultimately ineffective in addressing this kind of vulnerability—since even if they were entirely successful in undoing all the institutional equals of contemporary society, these regulations would not touch this fundamental vulnerability.

Pushing Lee’s argument further, we can say that this fundamental ontological fact of our linguistic vulnerability itself is posterior to a vulnerability to ideas or thoughts, even those left unexpressed. This is the cornerstone of Meir Dan-Cohen’s recent essay ‘Harmful Thoughts’. The salient aspect of Dan-Cohen’s complicated argument for us is his assertion that thoughts, of themselves, have extra-personal effects, even without the knowledge of the person affected and/or any immediate causal effect in their experiential frame. That is, a thought itself carries a kind of ontological volatility with it, which can have real-world impact on a person, even if they are ignorant of the content of the thought and the thought has no causal connection to an event in their life. Along with our linguistic vulnerability, we also have the ontological fact of our exposure to thoughts, beliefs, or attitudes which may never see the light of day.

The most significant manner in which thoughts can operate this way, Dan-Cohen claims, is as “the relational term that fixes a property of another person” (Dan-Cohen, p. 185). That is, sometimes, it is ideas themselves and not the person who holds them (with or without expressing them), which can determine a characteristic of another person. For example, if a person is concerned with being loved by a certain other person, they may become convinced that they are not so loved, despite the
affirmation of love by that person, because they believe that said other person’s “true thoughts” are not those they are expressing. In this case the idea “unloved by Mary” comes to be integral to Jane’s personhood, even though Mary might very well be busy trying by words and deeds to show Jane that this idea is false. Here we see that even though Jane does not know that Mary thinks “I do not love Jane” and even though this thought does not cause any actions by Jane which would lead Mary to that conclusion, the thought nevertheless affects—importantly: harms—Jane.

Such an example seems hopelessly trivial by comparison to the situation with hate speech, but the connection becomes clear when we name “unloved by Mary” an example of an “individual relational property,” and then think of what it means for some ideas to be “collective relational properties.” Rather than thinking of a certain Jane being affected by the thought (the “relational property”) “[I am] unloved by Mary,” here we might consider a particular person being similarly affected by the thought (the “relational property”) “[I am an] African-American.” The significance of this power of “collective relational properties” becomes clear when we bear in mind the importance “of the ever-present pragmatic background against which the judgments ascribing both properties and effects are made” (Dan-Cohen, p. 187). To continue with this example, this means that “African-American” (and collective relational properties more generally)—in a way considerably more pointed than “unloved by Mary” (and individual relational properties more generally)—do their work on the basis of a social reality that is by no means the active choosing of people living their lives in and through these ideas. More directly: being “African-American” also means any of a series of other identifying characteristics depending on the context in which it is a property one identifies with; it might mean, for example, “inferior,” “lazy,” “stupid,” “criminal.” In short, the thought itself might (depending on the pragmatic background in which a person holds it as a collective relational property) have all the positive, constructive content of an act of linguistic recognition like “You are part of a rich cultural tradition to which you might someday make a unique contribution” or, all the negative, destructive content of an act of linguistic harm like “You are a worthless nigger.” Since the thoughts themselves carry these varying significations, eliminating certain speech acts can in no way promise an alteration to the volatility of those thoughts; only a revision of the “pragmatic background” of language could begin to work in that direction. The psychological harm of “knowing” that you are seen through unjust stereotypes exists independently of ever actually hearing those stereotypes expressed to you in hateful speech. To get at that harm, and alleviate it in any way, it will not suffice to simply silence any public expression of the hateful calumny; we will have to address that “pragmatic background.”

B.

Having described the ontological character of our vulnerability to language and to thoughts, regardless of our direct awareness of them, it is valuable to see that harmfulness of certain kinds of language which circumscribe the possibilities of individuals. One such kind of speech is the hate speech that we have been discussing. Another, as Lee points out, is state speech which curtails the possibility individuals have to democratically deliberate on their identity. His example is racial classification. He points to a Louisiana case from the early 1980s in which a woman, Susie Phipps, who had lived her whole life as white, “challenged the right of the Louisiana State Bureau of Vital Statistics to classify her as ‘black’” (Lee, ‘Legal Weapons for the Weak?’, p. 884). While one might argue that there was less at stake in this for Phipps than for Plessy in 1896, what Lee sees as the underlying sameness is that in each case, the state has undermined the fundamental right of each
citizen to choose their identity. While Phipps may not have faced all the dehumanizing effects that Plessy had to, the fundamental harm—the intrinsic harm of the speech act (this time a speech act of the state, and one which is not precisely “hateful,” but which nevertheless shares the ontological force of words discussed above)—remains the same: the circumscription of individual possibilities at a psychological and ontological level. As Lee describes it this harm is primarily one of ontological possibility and only secondarily one of emotional suffering, a harm “that is primarily an injury to the capacity for self-definition and self-authentication and only secondarily an injury to the victim’s emotional status” (Lee, p. 886). This definition applies equally well to that hate speech of individuals as to the state speech of racial classification, as such demonstrates that we have met the test we were concerned with: describing a harm which does not rely on assuming that the speech was successful in causing pain in order to assert that there was in fact a harm.

Given the harmfulness—potential harmfulness—of thoughts, even if left unexpressed, one might expect Dan-Cohen to draw the conclusion that criminalizing them, or at least the speech that conveys them (in a way similar to the way we have criminalized reckless behavior), makes sense. He does not, for reasons which are both theoretically and practically compelling. Instead he argues that precisely because they do have this harmful potential we are best off keeping the discourse as open as possible. His view is that however true it may be that “the attitudes that hate speech expresses are, if pervasive, by themselves destructive of an aspect of their targets’ identity” it is also true that “the situation is only exacerbated by the indignity of illusion” (Dan-Cohen, p. 190). Because it is the case that all of the nastiness of a term like “nigger” can actually be imparted to someone, can thrive in their consciousness, without anyone ever expressing that nastiness in words, it actually turns out, on his account, that we only worsen our ontological vulnerability—and especially the vulnerability of those most undermined by hate speech—by trying to shove a cork in the mouths of those who might “air the nastiness out” by giving it voice. In this respect, insofar as it “brings things into the open,” hate speech—as opposed to the hateful thoughts it conveys themselves—performs a redeeming function as well.

However much this conclusion might seem a little glib—or even more than a little—it does seem to me clear that the ontological fact of our exposure to the harm hidden in the thoughts out there to think, whether or not those thoughts are ever actually expressed in speech makes it clear that regulation of such harm could never work. Even if you believe Dan-Cohen to be wrong about the “airing out” function of hate speech, it is incontrovertible that criminalizing this speech will not make our fundamental exposure to the psychological harms of hate go away. In other words, as Lee tells us, linguistic vulnerability is an ontological risk, it is an existential condition for the human being as a human being; as such, it is inseparable from our living in and through language, and therefore, it is, quite simply, ubiquitous. Therefore, the only thing that regulation could possibly hope to do is to quash our ability to express certain thoughts which are particularly heinous for our contemporary society and its norms. But all this could possibly achieve is a sort of “gag” function in one tiny part of our linguistic context that will force an eruption elsewhere, as Dan-Cohen argues.15

Therefore, it seems only right (if unfortunate) to conclude that the regulatory cost for criminalizing the “intrinsic” harm of hate speech is so great so as to conclude that it is strictly speaking impossible—both on ontological and socio-political grounds. On ontological grounds, it is impossible to regulate away our linguistic vulnerability; on socio-political grounds it is impossible
to regulate away the ideas that cause people to have (or expose themselves to) hateful thoughts which sometimes—but surely not always—lead to hateful speech.


A. These “intrinsic” harms often bleed (too often literally) into the “implicit” harms. Where the “intrinsic” harms are psychological and linguistic, bespeaking a fundamental fact of our existence at an ontological level; “implicit” harms are physical and contingent. These are much more simple to define and describe than the first sort of harm isolated: implicit harms are the physical harms either endorsed as having occurred in the past, or promised for the future, or both. If a person says to another person whom they know (or believe) to be homosexual, “That dirty faggot Matthew Shephard got what was coming to him,” that person is at the same time imparting (a) the intrinsic harm of the collective relational property “dirty faggot” and (b) the implicit harm of both Matthew Shepard’s physical suffering and some possible future physical suffering the object of the speech might be made to suffer. But if there is this relation between these kinds of harm, then what can we say about the second in itself?

At bottom, the idea of this implicit harm, is that not only is the speech itself violent insofar as it conveys, intrinsically, our fundamental linguistic vulnerability, and our inability to define our own identity by our standards, but it also is a fertilizer for further acts of violence--both linguistic and physical. As Minow says of hate speech acts: “The incidents also contribute to continuing dehumanization, enacting a victimization that can in a perverse way give new power to the picture of such group members as pathetic or vulnerable, and certainly new power to the sheer significance of such group memberships” (Minow, p. 34). We can understand the uniqueness of this second harm, therefore, to be constituted by the damage that is done to the object of a hate speech act insofar as they are identified with the victim of some past violence or as the victim of future physical violence.

When, for example, flyers are strewn across the Brown University campus saying that everything was much better when the brown, yellow, and black faces were in the kitchen, rather than the classroom, and that we ought to “keep white supremacy alive,” or when a fraternity at Arizona State University makes Jewish pledges read a statement that saying that their number is six million because that’s how many Jews died and “I should have been one of them” --when these events occur, there is not only the humiliation of that moment, which is a violence (a linguistic violence it is difficult to get precisely, but a violence nonetheless), but there is also the sustenance, the succour, the tilling of freshly dampened soil, given to and given for the possibility of future physical violence. Moreover, there is a quasi-physical violence, both after-image and pre-image of a more direct physical violence, actually gripping the object of the speech act.

B. The line between the intrinsic harm of a hate speech act on its own terms, and the implicit harm of the speech as a threat, can be a thin one. But this is no boon for the argument in favor of regulation. For, as we have seen, the intrinsic harm of hate speech is impossible to regulate away through criminalization. Therefore we are left with what harm lies in the implicit harmfulness independent
of the intrinsic harmfulness. But this, as we have seen, is nothing other the attempt to identify the
object of the speech with the object of some past violence, and/or to identify them as the object of
some future possible violence. Insofar as it is the latter, then we are talking about a threat,
harassment, or a “fighting words” situation, and these have already been criminalized—which is not
at all to say successfully prosecuted. Insofar as it is the former, then we can either say that (a) the
harm is in threatening a threat, in the sense of making a person anxious about the possibility that
this person speaking is trying to make them feel anxious, which amounts to the same as the
harmfulness of threats in general, and therefore cannot be criminally regulated any more than
statutes currently allow, or we can say that (b) the harm is some how undermining the personhood
of the person who is addressed by making them feel as though they themselves were the victim of
that former violence, at least for a moment. And if we go this second road, then we find ourselves
back in the same place where we were with harmful thoughts and the force of words: we simply
cannot eliminate that person’s exposure to this ontological diminution by illegalizing certain kinds
of speech acts.

This can be seen through one of the examples discussed. When someone forces—insofar as we can
see a strong sort of necessity in a fraternity hazing ritual—a young Jewish man to say “My number
is six million because that is how many died and I ought to have been one of them,” there is clearly
some harm being done beyond the intrinsic harm of exacerbating his exposure to linguistic
vulnerability. However, if we define that harm as threatening him with future violence, we see
immediately that there is no further criminal remedy peculiar to this phenomenon of hate speech.
And if we define it as a forced identification with the victims of the Nazi genocide, then we are left
with nothing more to “throw the book at” then the very kind of ontological circumcision which
we have already found criminalization to either be impotent against, or at best, to be a kind of
repression which pushes down on one geyser, only to have all that force spout out somewhere else.

In short, then, it turns out that regulating, or attempting to regulate, hate speech on the grounds of
redressing this second harm is impossible. For the “threatening” aspect of this implicit harm already
is criminal under harassment and/or stalking laws, and the “intrinsic” aspect has already been shown
to escape any attempted criminal regulation.

4. Conclusion: Striving to address, rather than failing to redress, the harms of hate speech.

At this point, it should be clear that I believe that there is a unique and discrete harm to be found in
these phenomena labeled “hate speech,” and that this harm ought to be condemned. The question of
course, immediately arises: it is all well and good to say that we must condemn these acts, but how
shall we condemn them? For example, we might equally wish to condemn a desperate, embarassed
parent who persistently and loudly verbally abuses their infant child who “won’t stop crying” while
sitting across from us on a bus or train and a parent who, in an act of similar but more extreme
desperate frustration, softly strangles their infant in their crib. I apologize for the awfulness of this
last image, but it makes the point quite sharply: the urgent need we feel to condemn each of these
persons is strong; the belief that prosecution is called for seems appropriate only to the latter.
Moreover, should we be speaking about the same parent in both cases, we would say that he or she
deserves prosecution, and that the earlier situation on the bus contributed to the climate which led to
the act deserving prosecution, but we would not for that reason call the first act “actionable.”

It turns out that condemnation as criminalization is both ineffective and normatively unconvincing.
I have argued above about this double-failure of criminalization as a response at a level of theoretical reflection, but a brief consideration of the experience of Canada’s regulation makes this all too clear as a matter of practicality. Canada lists two criminal offenses that capture the phenomena “hate speech.” The first, Section 318 of the Criminal Code, is a prohibition against advocating genocide; the second, Section 319, is a less definite prohibition against both the “public incitement of hatred” (section 1) and the “wilful promotion of hatred” (section 2). The punishments possible under these statutes are: (1) a prison sentence of no more than five years (for advocating genocide) or no more than two years (for either of the offenses described in Section 319); and (2) the forfeiture of any materials that were used to transmit the contact found criminal in the prosecution. Precisely given the nature of the harms outlined above, one has to ask what confiscating hateful literature and symbolic and short prison sentences really achieve in terms of ameliorating those harms. Moreover, whatever good one might see in those two outcomes, how much are they themselves mitigated by the self-righteousness that is sure to come from what will be seen as a “martyrdom” of those punished by the state in this way?

But the case against criminalization becomes even more damning when we look at the short of history of prosecution on these statutes. The first major case which tested the constitutional validity of these sections was R. v. Keegstra, in which Keegstra—a secondary school teach in Alberta, Canada—was prosecuted under Section 319. All the essential elements of the case fall very much in line with one’s intuitions. Keegstra is an anti-Semite who repeatedly and consciously uttered slurs against the Jews and spoke frequently about their nature, their practices, and their desire to rule the world. He stepped these practices up after the law was passed with precisely the intention of being prosecuted. He was successful. After six years, his initial conviction was upheld on appeal, and his sentence was commuted. And so, after all that time, his hateful language was amplified, his message spread, his personal time infringed on more by the media spotlight than any criminal punishment, and Canada had its judicial precedent for convicting future offenders through this Section of the Criminal Code.

The Canadian government tells this as a success story, but this seems more than dubious. Especially given that the decision which upheld the conviction demands an exceedingly difficult standard for the government in prosecuting on these statutes—the Crown must show that “the mental element is more than mere negligence, or recklessness as to the result”; that is, the government must show that the party consciously and explicitly intended to “incite hatred” and/or “wilfully promote hatred” and/or “advocate genocide”—it seems clear that the only time these statutes will actually be enforced is when attention-seeking hate-mongers violate the statute more or less expressly for the purpose of violating the statute. This suspicion is vindicated by one of the most recent applications of the statute, in which a vitriolic and ill-educated Toronto “preacher” (not affiliated with any established church or congregation) self-published a series of texts against Islam and Muslims, and received infinitely more attention through his arrest and prosecution than he ever had simply distributing his literature on street corners in the past.

Finally, there is the problem of just which groups ought to be listed as those it is criminal to incite (or wilfully promote) hatred about, and/or to advocate the genocide of. In the Canadian instance this initially included “any section of the public distinguished by colour, race, religion or ethnic origin,” but not (say) “gender, sexual orientation, or mental/physical ability.” Quite recently, the Canadian
Parliament added “sexual orientation” to the list of “identifiable groups” that it is criminal to
spread hate about or advocate the genocide of. This has had the (predictable) consequence of
making many religious people and organizations enemies of the legislation. And yet, not having
added the category would have made many other people and organization decry the legislation as
well. What this points to is not an accidental feature of the Canadian context, or a side problem that
could be addressed with better legislation and redefinition of the statute. Rather, it demonstrates that
criminalization only further exacerbates the deep divisions about core values that exist in pluralist
democracies, and that whatever (small) successes it might achieving in symbolic punishing those
that (by the definition of some) spread ideas which are intrinsically violent and therefore criminal, are
gained precisely at the expense of gratifying the (perhaps delusional) belief held by those who
endorse those ideas that they themselves are being victimized by the government. What little might
be achieved in redressing the root sources of the harms of hate speech is at least matched and
probably exceeded by the harm done in giving new fuel to the fire for those who are ready to speak or
happily hear those words in the first place.

Consider it granted that “regulation as criminalization” does not work. Given that there are distinct
harms here and they ought to be condemned, what condemnation is possible? I endorse a series of
iniatives: (1) targeting the extant, and acting against any future, local, state, or federal laws which
(like Minow’s twice-referenced California Proposition 187) give sustenance to hate-filled
ideologies; (2) symbolic acts of collective outrage and counter-speech (like the tiny Klan march in
New York which was tremendously outnumbered by protestors); (3) public repudiation and
individual outrage expressed by government officials and other opinion leaders, like President
Mbeki’s response to the “Kill the Boer!” chant controversy, which the South Africa Human Rights
Commission report referred to as a “public shaming”25; (4) educational programs for elementary
school students which respond to hate and show some of the root causes of it, as well as encourage
discussion and understanding; and (5) programs of research that “address potential connections
between intimate and intergroup violence” (cf. Minow, pp. 39, 47-50), the relationship between
violence in the house—be it spousal or child abuse—and the kind of violence under consideration
in Minow’s paper and mine.

Furthermore, as Lee argues, there are two necessary (though perhaps insufficient—even taken
together) alterations to democratic societies as sites of communicative action and of state speech
which will address the harms under discussion.26 First, rather than attempting to censor the speech
of (socially, politically, and economically advantaged) hate-mongers, we ought to increase access
and participation for those thus disadvantaged. This, he believes, can best be done through three
equalization processes: that is, equalizing the “performative capacity of disadvantaged speakers to”
(1) “undermine degraded meanings associated with the uncivil speech directed at them,” (2) “create
new meanings that have a fair chance at receiving public recognition,” and (3) “subject the speech
performances of socially advantaged speakers…to normative criticism” (Lee, p. 873). These
processes have the result of simultaneously minimizing the ability of advantaged speakers, and
maximizing the ability of disadvantaged speakers, to influence the discursively agreed upon
boundaries of “facts” about individuals as members of groups. Second, Lee tells us, we need to be
wary of the state’s power to undermine the right to expression held by individuals, understood as
the capacity to define oneself according to one’s own understanding of one’s identity. He uses the
Plessy case as an example of an individual’s expressive rights being undermined by state speech. In
this case, what is needed is a politicization and destabilization of the putatively natural categories of
race and ethnicity, which would have the result of simultaneously minimizing the state’s ability to, and maximizing the ability of disadvantaged speakers, to influence the discursively agreed upon boundaries of “facts” about individuals as members of groups. The net result of both of these changes is continuously, if only incrementally, empowering members of historically undermined groups without involving the state as a direct antagonist of the historically advantaged, with all the negative consequences described just above and throughout.

Beyond this, I do have two final suggestions. First, there is the possibility of using extant civil litigation. To the limited extent it is needful, symbolically and otherwise, to at least sometimes punish wrongdoers, these lawsuits can be much more effective at addressing those deeper sources of hatred. Minow gives the example of a “dramatically successful civil lawsuit brought against Tom and John Metzger and the White Aryan Resistance Organization for recruiting young skinheads who then killed an Ethiopian immigrant in Portland, Oregon” (Minow, p. 49). The idea was to name a whole bunch of actors who were involved in maintaining the conditions of possibility (the cable access show that recruited adolescents, the radio show used to do the same, and men who organized them) for the single act of violence perpetrated by a group of young men. What is interesting about this case for Minow is how it reaches so much deeper into the root causes of hateful violence than usual “bias crime” cases. It is because of this deeper reach—getting past this or that violent actor to the roots of the violence—that the impact of such a difficult, time-consuming, and important effort has ramifications far greater, Minow believes, for social justice than the banging-ones-head-against-the wall demanded by pursuing what might be, principally and practically, a dead end.

It seems at least in principle possible that this method of using civil litigation to expose the thick roots of hate-dissemination that lay beneath individual acts of hate—whether verbal or physical—and then to punish, in a pecuniary fashion those who consitute and give succour to those roots can be systematized in a satisfying way. The model I have in mind—though I have as yet failed to find any such model extant—would have the state cover the legal costs (that is to say the costs of counsel, as well as any court fees) of any party wishing to bring civil action against an accused perpetrator of hateful speech acts, once that party met certain limiting conditions. Should the party succeed in the action, the state’s costs would then be recovered from the defendant(s).

Second, apropos of my earlier example of child abuse in trying to distinguish instances where bad acts deserve both condemnation and prosecution from those that deserve condemnation and some other response than prosecution, I would like to suggest that there are various state actions in the register of “social service” that might be undertaken rather than state actions in the register of “criminal justice.” These two are linked, of course, and what I have in mind would probably have a legislative component, as there would have to be some mandate to be involved in some sort of “overcoming hate” workshops, tutorials, or trainings, the way that child abusers are held accountable by social services agencies entirely outside of court, once an offense has been claimed.

However empty, or to be more generous, thin and elusive, any of these--and even all of these--may seem in the face of the real, powerful, and awful reality of the persistence and prevalence of racist hate in the American polity, these suggestions all seem to me more plausible, more rationally defensible, and ultimately more likely to succeed than the prospect of regulating hate through the criminal justice mechanisms of the state.
In §1, surveying existing claims about the harm of hate speech, we found the claim at the heart of arguments made by those who want to see a criminal justice remedy for the harms of hate speech to be that the positive prescription to achieve greater equality needs be given greater heed, vis-à-vis the negative proscription of government’s ability to interfere with the speech acts of individuals. In the main body of the paper—in §§2 and 3, successively—we examined this claim in light of the proper understanding developed of the two main kind of harms that we can say to exist for victims of hate speech: namely, intrinsic harms (the psychological and ontological deficits and difficulties that are embodied in the speech itself and/or in the ideas it carries) and implicit harms (the future physical violence that the speech either promises or the former physical violence the speech endorses having occurred). We saw that in both cases the regulatory cost of trying to regulate the harm is so great as to be practically impossible. Thus, in present conclusion, we have tried to find a way to approach these harms differently, in a manner which recognizes both that (a) real harms are being perpetrated here and (b) the criminal justice system is not capable of redressing these harms. This “new approach” focuses on addressing the conditions for the possibility of hate speech, and its consequential harms, rather than failing to redress those harms once they occur.

Michael Weinman
New School for Social Research

Notes

1. Throughout the paper “regulation” shall be more or less synonymous with “criminalization.” Only in the conclusion will I point to the possibility of a kind of “regulatory response” which would not amount to a criminal justice mobilization and which might be more promising than “regulation as criminalization.”

2. There are also what I believe may be called the indirect harms of hate speech; namely, the structural and social deficits that hate speech neither engenders nor establishes, but does buttress and support. An example would be the indirect harms on homosexuals caused by the public audience and even institutional backing that someone like Rev. Fred Phelps gets when the judicial system finds that his proposed memorial for a public park in Casper, Wyoming (which reads: “Matthew Shephard entered Hell October 12, 1998, at age 21 in defiance of God’s solemn warning: Thou shalt not lie with mankind as with womankind; it is an abomination. Leviticus 18:22”). (See: http://firstamendmentcenter.org/news.aspx?id=12082 and http://firstamendmentcenter.org/news.aspx?id=12139). But a discussion of such harms belongs more properly to the analysis of those structural and social deficits themselves, and the way that institutions like courts perpetuate them, and less to a discussion of “hate speech.” All the same, I am convinced that the cost of regulating these harms is prohibitive, and thus a fuller discussion of them would not change the basic “anti-regulation as criminalization” argument I present.


4. Yet another way to try to describe those who have suffered the most persistently from discrimination and oppression. Here, in a discussion of the United States, the list must begin with
African-Americans, Hispanic Americans, Asian Americans, and then women, other immigrants, Jews, and at times Catholics. I think undermined is better than other choices because it shows that discrimination is always about dehumanization—about pulling out the subjectivity of certain people from under them; or, at least, trying to.


7. It is, of course, true that there are instances (such as sexual harassment laws) where a crime is said to have been committed independently of its success or lack thereof. However, the argument here is not so much whether someone has violated a law (defined this way or that), but rather whether or not on an ontological and psychological level a real harm has been inflicted. In the latter sense, the interpellation argument presumes what it must prove.


11. This argument can be found in Matsuda, ‘Public Response to Racist Speech’, pp. 31-38.

12. For those seeking an argument here, see Lee, ‘Legal Weapons for the Weak?’, 852-862, especially 858-862. Essentially, on jurisprudential grounds alone, the problem for the “pro-regulation” side here is that it is very difficult to ever assert any circumscription of a “fundamental” right (a “Bill of Rights” right, we could say) for any purpose, unless it is by another such “fundamental” right. So insofar as they are committed to agreeing that they are calling for a circumscription of speech rights, and they are defending this not on the basis of another “fundamental” right, their case is very hard.


15. This point may seem arguable, so let me first say that I am not saying that it is an absolute necessity that “gagging” certain extant forms of harmful speech will lead to further future harmful speech that has not yet been regulated. What I am saying, following Dan-Cohen, is that exposure to harm (in the form of the thoughts which may or may not be carried into the world in speech) is a fundamental aspect of the human condition, and that insofar as this linguistic vulnerability is a part of the unique harm of hate speech, it will never be adequately addressed by a criminal justice
response; which can never be more than repression. Is it a mistake that repression means both (a) an act by the state to hold something down or in check, and (b) an act by the individual consciousness to hide something from itself?


17. This example is loaded with more than just emotional urgency. It is not a mistake that I selected child abuse. Perhaps a reader would agree with my agreement with Butler and Minow that hate speech should not be regulated by criminal prosecution, but assert that regulation by the kind of social service monitoring and responses that child abuse warrants might be appropriate. I will take this up in the conclusion. Right here I only want to stress that I have not yet heard a compelling account for criminalizing child abuse in the register of the first example as a prosecutable infraction punishable by law. There are, of course, complications in the case of child abuse, such as (at least): (1) taking the child from the parent(s), and/or enforcing supervision, is or can be seen as “punishment”; (2) decisions about different state/social service actions (like removing the child from the parent(’s)(’s) custody) are often influenced by arguments regarding the effect that such actions would have on the victim, where the victim’s own sense of the offense is not always (or even often) seen to be a determining factor. All this shows just how complicated child abuse laws and other state and non-state responses to this phenomenon can be, but it also shows that there are acts which we strongly repudiate but do not seek to punish with jail time or monetary damages. I consider hate speech to be one of these, and to be (in many but surely not all respects) much like verbal child abuse.

18. See www.canlii.org/ca/sta/c-46/sec318.html and www.canlii.org/ca/sta/c-46/sec319.html.

19. And this is precisely the turn that political discourse seems to have taken in Canada in terms of the application of its hate speech regulations: at least four out of every five websites you find expressing a view on the matter takes up the view that the “swelling state” is infringing on the rights of either (a) individuals and/or (b) associations (such as churches). While this is by no means a scientific survey it does show that creating a regulation has mobilized anti-regulationists in Canada at least as much as not having created a regulation has mobilized pro-regulationists here.

20. A summary of the original case can be found at www.canlii.org/ca/cas/scc/1990/1990scc128.html; its decision (after appeal) can be found at www.canlii.org/ca/cas/scc/1996/1996scc19.html.


24. These being groups which the Justice ministry itself listed as “not included” (http://canada.justice.gc.ca).

25. Mbeki said: “nobody, whoever they are, has the right to call for the killing of farmers or Boers,
nor the right to threaten violence to advance their particular goals” (SAHRC, 13).

26. The harms of what we have called “hate speech,” which he calls “uncivil speech.”

27. The nature of these conditions would (obviously) have to be carefully determined, but might involve a kind of application and certification process which would be free to the party wishing to bring suit. Advising about the forms, about the litigation process, could be provided by volunteers from the legal community who would be recorded on a register kept by, say, city or county officials. The party wishing to bring suit would go to that county or city office, and then get the advice about how the procedure works. They would have no costs of their own to pay at any time.

28. I leave for another day a more precise determination of what such a program might look like.