Legal Theory, 7 (2001), 119–157. Printed in the United States of America
Copyright © Cambridge University Press 1352–3252/01 $9.50

MILLIAN PRINCIPLES, FREEDOM OF EXPRESSION, AND HATE SPEECH*

DAVID O. BRINK
University of California, San Diego

Hate speech employs discriminatory epithets to insult and stigmatize others on the basis of their race, gender, sexual orientation, or other forms of group membership. The regulation of hate speech is deservedly controversial, in part because debates over hate speech seem to have teased apart libertarian and egalitarian strands within the liberal tradition. In the civil rights movements of the 1960s, libertarian concerns with freedom of movement and association and equal opportunity pointed in the same direction as egalitarian concerns with eradicating racial discrimination and the social and economic inequalities that this discrimination maintained. But debates over hate speech regulation seem to force one to give priority to equality or to liberty. On the one hand, egalitarian concerns may seem to require restricting freedom of expression. Hate speech is an expression of discriminatory attitudes that have a long, ugly, and sometimes violent history. As such, hate speech is deeply offensive to its victims and socially divisive. Though one might well be reluctant to restrict speech, it might seem that the correct response to hate speech, as with other forms of discrimination, is regulation. On the other hand, libertarian concerns may seem to constrain the pursuit of equality. Though one may abhor hate speech and its effects, the cure might seem at least as bad as the disease. Freedoms of expression are among our most fundamental liberties. Offensive ideas are part of the price one must pay to protect these constitutional rights. This being so, it might seem that the correct response to hate speech is more speech—presumably egalitarian speech condemning hate speech—not the restriction of speech.

Hate speech regulation raises constitutional issues and issues of political

* An earlier version of this essay was presented as a Becker Lecture at Cornell University, at the University of Kansas, and at Macalester College. For helpful discussion of issues addressed here, I would like to thank Richard Arneson, Joshua Cohen, Carl Ginet, Donald Green, Jeff Greenburg, Tom Grey, Terence Irwin, Karen Jones, Philip Kitcher, Charles Lawrence, Julie Maybee, Richard Miller, Roosevelt Porter, Luke Robinson, Fred Schauer, Henry Shue, Kyle Stanford, Jason Stanley, Nicholas Sturgeon, Bonny E. Sweeney, Emilia Sweeney, Janet Swim, C.L. Ten, and Peter Vranas. Work on this essay was supported by a President's Research Fellowship in the Humanities from the University of California.

https://doi.org/10.1017/S1352325201072019 Published online by Cambridge University Press

morality. Though these two sorts of issues are not the same, they are difficult to separate entirely. I will discuss whether the regulation of hate speech can be justified as a matter of political morality, but I will also address the more obviously jurisprudential question of whether such regulation is consistent with sound First Amendment jurisprudence. I propose to approach these issues obliquely by asking what Millian principles concerning freedom of expression imply about the regulation of hate speech. John Stuart Mill's *On Liberty* offers the classic defense of freedom of expression and other liberties against governmental interference, one which finds important echoes in modern First Amendment jurisprudence.[1] If this is the case, exploring the implications of Millian principles for the regulation of hate speech should be part of forming sensible views about the political and constitutional legitimacy of such regulation.

Moreover, the implications of Millian principles for the regulation of hate speech are subtle in interesting and instructive ways. Mill's argument against censorship may seem to imply that all content-specific restrictions on speech, including the regulation of hate speech, are impermissible. It is true that Mill also accepts restrictions on liberty designed to prevent harm to others, and this might seem to make room for restrictions on speech that prohibit the use of offensive epithets. But Mill is very clear that mere offensiveness does not constitute harm (i 12; ii 44; iii 1; iv 3, 10, 12–21; v 5). Thus it is easy to see how one might think that Millian principles underwrite the libertarian view and condemn the regulation of hate speech. But Mill's position is more complicated than might at first appear. A proper understanding of his defense of freedom of expression and other liberties requires understanding the way in which he thinks certain liberties of thought, expression, and action are essential to the exercise of our higher capacities, in particular, our deliberative capacities.[2] But if our conception of fundamental liberties is governed by deliberative values, then the case against regulating hate speech is less clear, for it is arguable that hate speech does not advance but rather retards deliberative values. If so, then the grounds for caring about certain liberties of thought, expression, and action may allow and, indeed, require

1. Unless otherwise noted, references are to chapters and paragraphs in ON LIBERTY [*OL*] (E. Rapaport ed., Indianapolis: Hackett, 1978). I refer by chapter and paragraph to UTILITARIANISM [*U*] (G. Sher ed., Indianapolis: Hackett, 1979) and by both chapter and page number to *Considerations on Representative Government* [*CRG*] and *The Subjection of Women* [*SW*], both reprinted in THREE ESSAYS (Richard Wollheim ed., New York: Oxford University Press, 1975). Other references are to natural divisions in the text (if any) and page numbers in the editions in the COLLECTED WORKS OF JOHN STUART MILL (Toronto: University of Toronto Press). I will refer by book, chapter, and section and page number to *Principles of Political Economy* [*PPE*], COLLECTED WORKS, vols. ii-iii, and *A System of Logic* [*SL*], COLLECTED WORKS, vols. vii-viii.

2. Elsewhere I have explored this interpretive claim and its significance for understanding Mill's utilitarianism and the prospects for reconciling his utilitarianism and his recognition of rights to fundamental liberties; *see* David O. Brink, *Mill's Deliberative Utilitarianism*, 21 PHIL. & PUB. AFF. 67–103 (1992). My discussion of Millian principles here prescinds, so far as possible, from his specifically utilitarian commitments.

https://doi.org/10.1017/S1352325201072019 Published online by Cambridge University Press

the regulation of hate speech. If the full realization of deliberative values can occur only in a community regulated by mutual toleration and respect, then there will be an important sense in which the libertarian themes in Mill's version of liberalism are tempered or conditioned by egalitarian themes.

I. MILLIAN PRINCIPLES

In attempting to understand Mill's views about freedom of expression, it can help to remember their role in his more general defense of individual liberties in *On Liberty*. In an early and famous passage, Mill offers one formulation of his basic principles:

> The object of this essay is to assert one very simple principle, as entitled to govern absolutely the dealings of society with the individual in the way of compulsion and control, whether the means used be physical force in the form of legal penalties or the moral coercion of public opinion. That principle is that the sole end for which mankind are warranted, individually or collectively, in interfering with the liberty of action of any of their number is self-protection. That the only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others. His own good, either physical or moral, is not a sufficient warrant. He cannot rightfully be compelled to do or forbear because it will be better for him to do so, because it will make him happier, because, in the opinions of others, to do so would be wise or even right. These are good reasons for remonstrating with him, or reasoning with him, or persuading him, or entreating him, but not for compelling him or visiting him with any evil in case he do otherwise. To justify that, the conduct from which it is desired to deter him must be calculated to produce evil to someone else. The only part of the conduct of anyone for which he is amenable to society is that which concerns others. In the part which merely concerns himself, his independence, is, of right, absolute. Over himself, over his own body and mind, the individual is sovereign. (i 9)

In this passage, Mill distinguishes paternalistic and moralistic restrictions of liberty from restrictions of liberty based upon the harm principle. A's restriction of B's liberty is paternalistic if it is done for B's own benefit; it is moralistic if it is done to ensure that B acts morally or not immorally. By contrast, A's restriction of B's liberty is an application of the harm principle if A restricts B's liberty in order to prevent harm to someone other than B. Here Mill seems to say that a restriction on someone's liberty is legitimate if and only if it satisfies the harm principle (cf. iv 1–4, 6; v 2). Exactly what, in Mill's view, will count as a harm for purposes of the harm principle is complicated. He clearly denies that mere inconvenience or offense is a harm. Rather, in order to satisfy the harm principle, an action must actually violate or threaten imminent violation of those important interests of oth-

https://doi.org/10.1017/S1352325201072019 Published online by Cambridge University Press

ers in which they have a right (i 12; iii 1; iv 3, 10, 12; v 5).[3] Familiar provisions of the criminal law, including laws against murder, rape, and assault, presumably satisfy the harm principle. Mill appears to think that restrictions of liberty based upon the harm principle are unproblematic (*but see* v 3). By contrast, he sometimes claims, as he does here, that paternalistic and moralistic restrictions of liberty are never justified (*cf.* iv 3, 4, 6; v 2).

Mill thinks that these principles provide a wide-ranging defense of individual liberties against governmental interference. To defend these commitments, Mill turns to freedom of expression. He believes that there is general agreement on the importance of free speech and that, once the grounds for free speech are understood, this agreement can be exploited to support a more general defense of individual liberties (i 16; iii 1).

II. MILL AGAINST CENSORSHIP

Mill's discussion of censorship in chapter ii focuses on censorship whose aim is to suppress false or immoral opinion (ii 1–2).[4] He mentions four reasons for maintaining free speech and opposing censorship:

1. A censored opinion might be true (ii 1–20, 41).
2. Even if literally false, a censored opinion might contain part of the truth (ii 34–39, 42).
3. Even if wholly false, a censored opinion would prevent true opinions from becoming dogma (ii 1–2, 6, 7, 22–23, 43).
4. As a dogma, an unchallenged opinion will lose its meaning (ii 26, 43).

The first two claims represent freedom of expression as instrumentally valuable; it is valuable, not in itself, but as the most reliable means of producing something else that Mill assumes is valuable (either extrinsically or intrinsically), namely, true belief. Of course, the most reliable means of promoting true belief would be to believe everything. But that would bring a great deal of false belief along too. A more plausible goal to promote would be something like the ratio of true belief to false belief.[5] Freedom of expression might then be defended as a more reliable policy for promoting

3. For a useful discussion, within a largely Millian perspective, of the difference between harm and offense, *see* JOEL FEINBERG, THE MORAL LIMITS OF THE CRIMINAL LAW, vol. 1, HARM TO OTHERS 31–36, 45–51 (New York: Oxford University Press, 1984).

4. Mill seems not especially concerned with a kind of censorship that often concerns us, namely, censorship whose aim is to suppress *true*, rather than false, beliefs thought by censors to be especially dangerous, such as certain kinds of classified information. Presumably, he could and would claim that such censorship is permissible just in case it satisfies the harm principle.

5. Even this version of the goal is probably too thin; it's hard to believe that it ought to be one's goal to promote the ratio of true belief to false belief, independently of the significance or importance of the beliefs in question.

https://doi.org/10.1017/S1352325201072019 Published online by Cambridge University Press

the ratio of true belief to false belief than a policy of censorship. This rationale for freedom of expression is echoed by Justice Holmes in his famous dissent in Abrams v. United States when he claims that the best test of truth is free trade in the marketplace of ideas.[6]

But this rationale for freedom of expression is pretty weak. It might justify freedom of expression in preference to a policy of censorship whenever the censor thinks an opinion false. But it doesn't justify freedom of expression in preference to more conservative forms of censorship. If the question is what policies are likely to increase the ratio of true belief to false belief, it would seem that we should employ conservative criteria of censorship and censor those opinions for whose falsity there is especially clear evidence.[7] We would be on good ground in censoring flat-earthers.[8] In any case, if, even if only contrary to fact, we had extremely knowledgeable and reliable censors who censored all and only false beliefs, this rationale for freedom of expression would provide no argument against censorship.

However, Mill also suggests that freedom of expression is needed to keep true beliefs from becoming dogmatic. In this suggestion, I think, lie the resources for a more robust defense of freedom of expression, in part because it is intended to rebut the case for censorship even on the assumption that all and only false beliefs would be censored (ii 2, 21). Mill's argument is that freedoms of thought and discussion are necessary conditions for fulfilling our natures as progressive beings (ii 20). To understand this argument, we need to understand the evaluative assumptions on which it rests.

In *On Liberty*, Mill claims that his defense of liberty relies on claims about the happiness of people as progressive beings:

> It is proper to state that I forego any advantage which could be derived to my argument from the idea of abstract right as a thing independent of utility. I regard utility as the ultimate appeal on all ethical questions; but it must be utility in the largest sense, grounded on the permanent interests of man as a progressive being. (i 11)

Mill thinks that it is our deliberative capacities, especially our capacities for practical deliberation, that mark us as progressive creatures and that, as a result, the principal ingredient of our happiness or well-being must exercise these deliberative capacities. At its most general, practical deliberation involves reflective decision-making. In *On Liberty* Mill thinks of practical

6. 250 U.S. 616 (1919) (upholding the conviction of a wartime pamphleteer on behalf of the Russian revolution under the Espionage Act of 1917).

7. So, contrary to what Mill sometimes assumes (ii 3, 41), the argument for censorship need not presuppose the infallibility of the censor. Fortunately, Mill's best arguments against censorship do not require this assumption about the argument for censorship.

8. *Cf.* Alvin Goldman, *Epistemic Paternalism: Communication Control in Law and Society*, 88 J. PHIL. 113–131 (1991).

https://doi.org/10.1017/S1352325201072019 Published online by Cambridge University Press

deliberation in terms of capacities to form, assess, choose, and implement projects and goals:

> He who lets the world, or his own portion of it, choose his plan of life for him has no need of any other faculty than the ape-like one of imitation. He who chooses his plan for himself employs all his faculties. He must use observation to see, reasoning and judgment to foresee, activity to gather materials for decision, discrimination to decide, and when he has decided, firmness and self-control to hold his deliberate decision. And these qualities he requires and exercises exactly in proportion as the part of his conduct which he determines according to his own judgment and feelings is a large one. It is possible that he might be guided in some good path, and kept out of harm's way, without any of these things. But what will be his comparative worth as a human being? (iii 4)

Mill makes similar claims about the role of deliberative capacities in the happiness of progressive beings in his discussion of higher pleasures in chapter ii of *Utilitarianism;* there he contrasts the examined life Socrates led and the life of a contented swine, and accords the former incomparably greater value (*U:* ii 6).[9]

Even if we agree that these deliberative capacities are unique to humans or that humans possess them to a higher degree than other creatures, we might wonder in what way their possession marks us as progressive beings or how their exercise is important to human happiness. Mill thinks an account of human happiness ought to reflect the kinds of beings we are or what is valuable about human nature. Though he is not as clear about this as one might like, his discussion of responsibility in *A System of Logic* ("Of Liberty and Necessity") suggests that he thinks that humans are responsible agents and that this is what marks us as progressive beings. There he claims that capacities for practical deliberation are necessary for responsibility. In particular, he claims that moral responsibility involves a kind of self-mastery or self-governance in which one can distinguish between the strength of one's desires and their suitability or authority and in which one's actions reflect one's deliberations about what is suitable or right to do (*SL:* VI, ii, 3). Non-responsible agents, such as brutes or small children, appear to act on their strongest desires or, if they deliberate, to deliberate only about the instrumental means to the satisfaction of their strongest desires. By contrast, responsible agents must be able to deliberate about the appropriate-

9. A deliberative conception is reflected in claims Mill makes elsewhere. In *Considerations on Representative Government* he claims that a principal aim of government is the improvement of its citizens, conceived as the development of their intellectual, deliberative, and moral capacities (*CRG:* esp. chs. ii-iii). In *The Subjection of Women* he explains the unhappiness for women in their subjection in terms of the way sexist institutions and attitudes prevent them from developing their rational and deliberative powers (*SW:* iv 542–8). In various places Mill also expresses reservations about charities that encourage dependence of the beneficiary on her benefactor and so undermine the beneficiary's self-development and self-respect (*SW:* iv 532; *PPE:* V.xi.13 960–962).

https://doi.org/10.1017/S1352325201072019 Published online by Cambridge University Press

ness of their desires and regulate their actions according to these deliberations. If this is right, then Mill can claim that possession and use of our deliberative capacities mark us as progressive beings and constitute the principal ingredient in human happiness.

This puts us in a position to explain Mill's claim that the value of freedom of expression lies in keeping true beliefs from becoming dogmatic. This claim reflects Mill's view that freedoms of thought and discussion are necessary conditions for fulfilling our natures as progressive beings (ii 20). We can see Mill appealing to a familiar distinction between *true belief*, on the one hand, and *knowledge* or *justified true belief*, on the other hand. Progressive beings seek knowledge or justified true belief, and not simply true belief. Whereas the mere possession of true beliefs need not exercise one's deliberative capacities, because they might be the product of indoctrination, their justification would; one exercises deliberative capacities in the justification of one's beliefs and actions that is required for theoretical and practical knowledge.[10] This is because justification involves comparison of and deliberation among alternatives (ii 6, 7, 8, 22–23, 43). Freedoms of thought and discussion are essential to the justification of one's beliefs and actions, because individuals are not cognitively self-sufficient (ii 38, 39; iii 1). Sharing thought and discussion with others, especially about important matters, improves one's deliberations. It enlarges the menu of options, by identifying new options worth consideration, and helps one better assess the merits of these options, by forcing on one's attention new considerations and arguments about the comparative merits of the options. In these ways, open and vigorous discussion with diverse interlocutors improves the quality of one's deliberations. This being so, censorship, even of false belief, can rob both those whose speech is suppressed and their audience of resources that they need to justify their beliefs and actions (ii 1).

It is important not to overstate the significance of this argument against censorship. Deliberative values do not always speak in favor of expanding one's option set. Cognitively limited agents cannot consider all logically possible options, and careful consideration of many options—especially irrelevant options and options known to have failed—is likely to retard, rather than advance, their deliberations. More options are not always better than fewer.[11] It is an important but difficult project, well beyond the scope of this essay, to develop principles by which to structure the alternative space of cognitively limited beings and by which they should regulate their search of this space. It is enough for our purposes, I think, to note that Mill's appeal to deliberative values explains why it is often wrong to censor even false beliefs without implying that censorship is always wrong.

10. *Cf.* T.M. Scanlon, *A Theory of Freedom of Expression*, 1 PHIL. & PUB. AFF. 204–26 (1972) and C.L. TEN, MILL ON LIBERTY 126–28 (New York: Oxford University Press, 1980).

11. *Cf.* Gerald Dworkin, *Is More Choice Better than Less? in* THE THEORY AND PRACTICE OF AUTONOMY (New York: Cambridge University Press, 1988).

https://doi.org/10.1017/S1352325201072019 Published online by Cambridge University Press

III. MILL'S GENERAL DELIBERATIVE ARGUMENT

Though important in its own right, Mill's defense of freedom of thought and discussion provides the resources for an argument for various basic liberties; the deliberative rationale for freedoms of thought and discussion is a special case of a more general defense of basic liberties of thought and action that Mill offers in chapter iii of *On Liberty*. A good human life is one that exercises one's higher capacities (i 11; ii 20; iii 1–10); a person's higher capacities include her deliberative capacities, in particular, capacities to form, revise, assess, select, and implement her own plan of life.[12] This kind of self-government requires both positive and negative conditions. Among the positive conditions it requires is an education that develops deliberative competence (v 12–15). Among the negative conditions that self-government requires are various liberties of thought and action. If the choice and pursuit of projects and plans are to be deliberate, they must be informed as to the alternatives and their grounds, and this requires intellectual freedoms of speech, association, and press. If there are to be choice and implementation of choices, there must be liberties of action such as freedom of association, freedom of worship, and freedom to choose one's occupation. Indeed, liberties of thought and action are importantly related. Mill values diversity and experimentation in lifestyles not only insofar as they are expressions of self-government but also insofar as they enhance self-government. For experimentation and diversity of lifestyle expand the deliberative menu and bring out more clearly the nature and merits of options on the menu (ii 23, 38; iii 1).[13] But diversity and experimentation presuppose liberties of action, and in this way liberties of action, as well as thought and discussion, are essential to the full exercise of deliberative capacities.

This interpretation provides Mill with a robust rationale for various liberties of thought and action; they are important as necessary conditions for exercising our deliberative capacities and so for producing the chief

12. The value of deliberative capacities, within Mill's brand of perfectionism, certainly provides the basis for criticizing some lives as shallow and undemanding, even when these lives are contented and successful in their own terms. But because capacities for practical deliberation can be realized and expressed equally or incomparably well in many different kinds of lifestyles, Mill can and does recognize very diverse kinds of valuable lives. We might say that Mill's brand of perfectionism respects *moderate pluralism* about the good, even though it rejects *content-neutrality* about the good. *See* Brink, *Mill's Deliberative Utilitarianism*, 79–80.

13. Indeed, Mill suggests that where insufficient diversity of opinion exists naturally, it is essential to take various steps to increase such diversity (ii 23). His defense of proportional representation (*CRG:* vii 260–62) and state support for the arts (*PPE*: V.xi.15 968–70) should be seen as institutional mechanisms designed to increase the diversity of activities and voices so as to enhance public and private deliberation. Moreover, we can see how such Millian arguments might support adoption of a fairness in broadcasting doctrine that required media to devote broadcast time to political and educational issues and to allocate time to represent diverse viewpoints on the subjects covered; *cf.* Red Lion Broadcasting Co. v. FCC, 395 U.S. 367 (1969) (upholding the FCC's fairness doctrine and, in particular, the personal attack rule against First Amendment challenge).

https://doi.org/10.1017/S1352325201072019 Published online by Cambridge University Press

ingredients of human happiness.[14] But it is also important to see that Mill is not endorsing an unqualified right to liberty. For one thing, he can distinguish the importance of different liberties in terms of their role in practical deliberation. Some liberties are *more central* to the exercise of deliberative capacities than others. It seems plausible that liberties of speech, association, worship, and choice of profession are more important than liberties not to wear seat belts or to dispose of one's gross income as one pleases, because restrictions on the former interfere more than restrictions on the latter with deliberations about what sort of person to be. If so, Millian principles, properly understood, defend rights to *basic liberties*, rather than a right to liberty per se.[15] Moreover, even the exercise of basic liberties is limited by the harm principle, which justifies restricting liberty to engage in actions that cause harm or threaten imminent harm to others.

IV. DELIBERATIVE VALUES AND FIRST AMENDMENT CATEGORIES

To see what Millian principles imply about the moral and constitutional permissibility of regulating hate speech, it may help to understand the role that deliberative values play in shaping parts of the constitutional background to debates about hate speech.

When the Supreme Court determines that a particular interest or liberty is a fundamental constitutional value, it accords that value special protection by subjecting legislation that interferes with that value to *strict scrutiny* or some comparable standard. To pass strict scrutiny, legislation must pursue a compelling state interest in the least restrictive manner possible. Strict

14. To account for the robust character of the argument, it is tempting to suppose that Mill thinks these basic liberties are themselves important intrinsic goods. But limitations in the scope of Mill's argument show that this cannot be right. In the introductory chapter, Mill insists that his defense of liberties applies only to those with mature deliberative faculties (i 10). This restriction makes no sense if basic liberties are intrinsic goods, for then it should always be valuable to accord people liberties—a claim that Mill here denies. It makes perfect sense if the liberties in question, though not intrinsically valuable, are necessary conditions for realizing important intrinsic goods, for then there need be no value to liberty where other necessary conditions for the realization of these goods—in particular, mature deliberative capacities—are absent.

15. This conclusion is overdetermined insofar as Mill also recognizes permissible restrictions on various individual liberties. He limits the application of his principles to those of mature deliberative competence (i 10). Also, he qualifies his blanket prohibition on paternalism to claim that no one should be free to sell herself into slavery, and this exception, he claims, is just an extreme example of a more common phenomenon (v 11). Furthermore, in *Principles of Political Economy* and *Considerations on Representative Government*, Mill endorses various kinds of legislation enacted pursuant to the community's interest—redistributive taxation; Poor Laws; labor regulation; provision for a common defense, public education, and community infrastructure; and state support for the arts—some of which would be difficult to justify in terms of the harm principle. It's difficult to see how Mill could recognize all these restrictions on individual liberties if he recognizes a libertarian right to liberty as such rather than to specific basic liberties.

https://doi.org/10.1017/S1352325201072019 Published online by Cambridge University Press

scrutiny and its relatives contrast with a weaker standard of review, known as *rational basis review,* that is applied to legislation affecting interests and liberties that are not fundamental. To pass rational basis review, legislation need only pursue a legitimate interest in a reasonable manner. With some notable exceptions in which the Court recognizes intermediate levels of scrutiny, its analysis of the importance of interests or liberties and associated standards of scrutiny is generally *bivalent:* interests or liberties are either fundamental or they are not; fundamental ones trigger strict scrutiny or some comparable standard, whereas non-fundamental ones trigger rational basis review or some comparable standard.[16] For the most part, liberties of expression are treated as fundamental liberties, because of the central role open discussion plays in both public and private deliberations. Insofar as liberties of expression are fundamental, the Court protects them by subjecting legislation that interferes with them to strict scrutiny or some comparably exacting standard, such as the clear and present danger test.[17]

But not all liberties of expression are treated the same. For instance, First Amendment analysis distinguishes between *content-neutral* restrictions on speech that restrict the time, manner, and place of speech but not its content, and *content-specific* restrictions that restrict some forms of speech on account of the topic discussed or the viewpoint expressed in the speech. Whereas content-specific restrictions are subject to heightened scrutiny, content-neutral restrictions are subject to weaker forms of scrutiny.[18] Deliberative values would seem to explain the Court's special concern with content-specific restrictions. Often, time, manner, and place restrictions leave open many avenues of expression and so do not significantly restrict the production, distribution, or consumption of ideas. By contrast, content-specific, especially viewpoint-specific, restrictions make it harder for certain messages to be heard and evaluated. If the representation of diverse perspectives, even mistaken ideas, improves public and private deliberations,

16. The Court's treatments of commercial speech, under First Amendment jurisprudence, and gender classifications, under Equal Protection jurisprudence, are among the exceptions to this rule, insofar as the Court subjects restrictions on commercial speech and regulations distributing social benefits and burdens by gender to an intermediate standard of review. I discuss the adequacy of the bivalence assumption in Section VII.

17. Besides employing strict scrutiny, there is an important line of cases articulating a "clear and present danger" standard. *See, e.g.,* Schenck v. United States, 249 U.S. 47 (1919) (upholding conspiracy convictions, under the Espionage Act of 1917, for the distribution of literature aiming to obstruct the military draft), Terminiello v. Chicago, 337 U.S. 1 (1949) (invalidating a disorderly conduct conviction in which the jury was instructed that it could convict if it found that the speech in question "stirs the public to anger, invites dispute, brings about a condition of unrest, or creates a disturbance"), and Brandenburg v. Ohio, 395 U.S. 444 (1969) (striking down convictions of the organizers of a KKK rally under the Ohio Criminal Syndicalism Act). I take the "clear and present danger" standard, as currently understood, to be more or less a special case of strict scrutiny.

18. *See, e.g.,* Geoffrey Stone, *Content Regulation and the First Amendment,* 25 WM. & MARY L. REV. 189–252 (1983). However, content-neutral restrictions that have a profound or disparate impact on speech are highly problematic.

https://doi.org/10.1017/S1352325201072019 Published online by Cambridge University Press

then there is general reason to think that content-specific restrictions constrain deliberative values in unacceptable ways.[19]

But not all content-specific regulations are thought to restrict fundamental liberties. First Amendment jurisprudence also distinguishes between *low-value* and *high-value* speech. The liberty to engage in low-value speech is not a fundamental liberty; content-specific regulation of low-value speech, as a result, need not satisfy strict scrutiny. By contrast, other forms of speech are high-value, and the liberty to engage in them is a fundamental liberty; as a result, content-specific regulation of high-value speech must satisfy strict scrutiny or some comparable standard. The Court formulated the distinction between low-value and high-value speech in Chaplinsky v. New Hampshire:

> There are certain well-defined and narrowly limited classes of speech, the prevention and punishment of which have never been thought to raise any Constitutional problem. These include the lewd and the obscene, the profane, the libelous, and the insulting or "fighting" words—those which by their very utterance inflict injury or tend to incite an immediate breach of the peace. It has been well observed that such utterances are no essential part of any exposition of ideas, and are of such slight social value as a step to truth that any benefit that may be derived from them is clearly outweighed by the social interest in order and morality.[20]

19. Others have stressed the important role that vigorous and open discussion plays in democratic political deliberations. *See, e.g,* ALEXANDER MEIKELJOHN, FREE SPEECH AND ITS RELATION TO SELF-GOVERNMENT (New York: Harper and Row, 1948) and CASS SUNSTEIN, DEMOCRACY AND THE PROBLEM OF FREE SPEECH (New York: Free Press, 1993). This is important, but only part of the story. Though many kinds of speech may be relevant, directly or indirectly, to collective self-governance, I see no reason to suppose that this sort of political relevance is necessary for speech to count as high-value. As Mill's arguments show, liberties of expression are important for individual as well as collective self-governance—for personal as well as public deliberation. If we recognize the role of liberties of thought and discussion within moral as well as political agency, we can plausibly explain why fundamental liberties of thought and expression are not restricted to political speech. For a more multi-dimensional rationale for expressive liberties, *see* Joshua Cohen, *Freedom of Expression*, 22 PHIL. & PUB. AFF. 207–63 (1993). Cohen invokes expressive and informational as well as deliberative interests in freedom of expression. However, it is arguable that both expressive and informational interests can and should be fitted under the umbrella of deliberative values. Insofar as the exercise of deliberative capacities involves decision-making that is informed as to the alternatives, informational interests must figure prominently within an account of deliberative values. Moreover, as Mill makes clear, deliberative values are at stake not just in decisions about what to do but also in decisions about *how* to do it and in the actual implementation of projects and plans (iii 4). If so, determination of one's mode of expression is an important deliberative value. Moreover, the expression of one's deliberations in choices and actions that are accessible to others expands the deliberative menu available to all and brings out more clearly the nature and merits of options on the menu (ii 23, 38; iii 1). In these ways, expressive interests are also essential to deliberative values. If so, we may not need a more multi-dimensional rationale for expressive liberties than that provided by appeal to deliberative values in order to recognize the importance of informational and expressive interests.

20. 315 U.S. 568 (1942) (upholding a state prohibition on the use of offensive language in face-to-face exchanges in public spaces) at 571–572.

https://doi.org/10.1017/S1352325201072019 Published online by Cambridge University Press

Here the Court associates central First Amendment liberties with what is an essential part of the exposition of ideas and what is of value as a step toward truth. Like Mill, the Court justifies freedom of expression as a way of promoting true belief. However, if the Court values freedom of expression only as a means of promoting true belief, then it becomes difficult to extend protection to false beliefs, as the Court has. But we need not interpret the Court as valuing freedom of expression only as a means of acquiring true beliefs. The Court appeals to what is an essential part of the exposition of ideas and what is of value as a step toward truth. We can see this rationale as invoking, as Mill also does, deliberative values about the way in which beliefs are formed, assessed, and accepted. If we interpret the Court's rationale this way, we can provide a more wide-ranging conception of high-value speech that includes the advocacy of some false beliefs and recognizes some forms of low-value speech. On this reading, low-value speech is not protected, because it does not contribute to the deliberative values that justify protecting other forms of speech.

This claim can be made out reasonably well in the case of libel. Libel is false and defamatory speech in which the speaker knows that her statement is false and defamatory or acts in reckless disregard of these matters.[21] It is true, as Mill claims, that the careful consideration of claims, advanced in good conscience, that are in fact false can advance deliberation by forcing us to consider the grounds of their falsity. But libelous speech is not advanced in good conscience. It is arguably a case in which more speech is not better insofar as the introduction of false and harmful claims with no concern for their truth and consequences arguably hinders, rather than promotes, reasoned assessment of issues.

Deliberative values also explain why fighting words are low-value speech. *Chaplinsky* characterizes fighting words as those that "by their very utterance inflict injury or tend to incite an immediate breach of the peace." As a matter of subsequent constitutional doctrine, the Court has interpreted the category of fighting words narrowly, focusing on their tendency to incite violence.[22] Fighting words, so understood, are words that in their context tend to evoke visceral and violent—rather than articulate—responses. However, it would be a mistake to focus on pugilistic responses, and this is why

21. *See* RESTATEMENT (SECOND) OF TORTS §§ 558, 559, 568 (1977). Actually, this is an account of the libel of a public figure. *See* RESTATEMENT § 580A and New York Times v. Sullivan, 376 U.S. 254 (1964) (preventing a public official from recovering damages for a defamatory falsehood relating to his official conduct unless he can prove that the statement was made with malice—that is, with knowledge that it was false or with reckless disregard to its truth). It is somewhat easier to establish libel in cases in which the speech concerns private persons; whereas a private individual need only establish negligence on the speaker's part, a public figure must establish at least reckless disregard. *See* RESTATEMENT § 580B.

22. *See, e.g,* Gooding v. Wilson, 405 U.S. 518 (1972) (overturning convictions under a Georgia law that prohibited speech directed at another that employs abusive language "tending to cause a breach of the peace" on the ground that as interpreted by state courts it was not in fact restricted, as in *Chaplinsky,* to speech that tends to incite an immediate breach of the peace).

https://doi.org/10.1017/S1352325201072019 Published online by Cambridge University Press

Chaplinsky rightly construes fighting words more broadly, so as to include words whose utterance would cause injury in a reasonable person. A natural response to the use of insulting epithets in many such contexts is visceral but non-violent; the victim of fighting words might be intimidated and silenced as well as provoked. Whether silence or fisticuffs, the natural response is not articulate. But then, fighting words simply express, without articulating, the speaker's perspective, and they invite various inarticulate responses. If so, we can see why the Court might reasonably claim that they do not contribute to deliberative values but often hinder them.

V. HATE SPEECH

To see what deliberative values imply about the permissibility of regulating hate speech, we need to know a little more about hate speech and its regulation. Hate speech should be distinguished from related phenomena. Whereas *hate crimes* are forms of criminal misbehavior that are motivated by prejudice or bias, and *bias-motivated sentence enhancements* impose more severe penalties for crimes when they are motivated by bias, *hate speech* is expression that vilifies or harasses on the basis of the target's race, gender, sexual orientation, or other forms of group membership. Although the state might decide to criminalize hate speech, hate speech might be regulated without invoking the criminal law. It might be recognized as a tort in civil law, and various institutions and associations might prohibit their members from engaging in hate speech, imposing various kinds of sanctions for noncompliance. For the most part, I shall focus on the legitimacy of this last form of hate speech regulation.

Hate speech regulations of one form or another are quite common elsewhere in the international community. In 1969 the United Nations adopted the International Convention on the Elimination of All Forms of Racial Discrimination. Article 4 of that Convention requires, in part, that member-states "declare as an offense punishable by law all dissemination of ideas based on racial superiority or hatred [and] incitement to racial discrimination."[23] Similar, though sometimes narrower, hate speech regulations are part of domestic law in Canada, Great Britain, Germany, France, Italy, Finland, Norway, and the Netherlands, among other nations.[24] Within the United States, however, matters are different. Whereas many states have enacted hate crime statutes and bias-motivated crime sentence enhancements,[25] hate speech regulations have been less common and have been

23. As reported in *Positive Measures Designed to Eradicate All Incitement to, or Acts of, Racial Discrimination: Report of the U.N. Committee on the Elimination of Racial Discrimination*, U.N. GAOR, 27th Sess., U.N. Doc. A/CONF.119/10 (1986).

24. *Ibid.*; *cf.* David Kretzmer, *Freedom of Speech and Racism*, 8 CARDOZO L. REV. 445–513 (1987).

25. *See, e.g.*, Wisconsin v. Mitchell, 113 S. Ct. 2194 (1993) (upholding conviction under Wisconsin's hate-crime penalty enhancement statute on the ground that the statue concerns conduct unprotected by the First Amendment).

https://doi.org/10.1017/S1352325201072019 Published online by Cambridge University Press

treated as constitutionally problematic.[26] Most of the attempts to regulate hate speech have involved campus speech codes adopted by public and private institutions of higher education to protect members of historically marginalized groups on campus from offensive and discriminatory harassment by other members of the campus community. In many instances, hate speech codes were adopted at least partly in response to specific campus incidents.

It is useful to have a sense of the nature of some of these incidents. Some of the incidents involved racist speech.[27] At the University of Wisconsin, fraternity members staged a mock slave auction with white pledges in blackface. In a separate incident at the University of Wisconsin, white male students followed black females students across campus shouting "I've never tried a nigger before." At Purdue University, a counselor found the words "Death Nigger" scratched on her door. At Smith College an African student found the following message under her door "African Nigger do you want some bananas? Go back to the jungle." And at Stanford University, in response to campus debates about the Western civilization curriculum, two white students defaced a poster of Beethoven to portray a stereotypical black face and displayed it in a predominantly black dormitory. Some other incidents involved homophobic speech at Yale University.[28] In one case, lesbian women were stalked and called "dykes" and "disgusting bitches." According to another lesbian, she was accosted by five men who "told me they wanted me to give them a blow-job and to fuck me. They asked me if I was dyke and repeated the question when I ignored them. They said they wanted to teach me about sex." And a gay man reported receiving obscene phone calls and having obscene insults left on his memo board. Such incidents are all too common; a disturbingly high percentage of members of traditional target groups report experiencing hate speech and other forms of discriminatory harassment on campus.[29]

26. *See* James B. Jacobs and Kimberly Potter, *Hate Crimes: A Critical Perspective*, *in* CRIME AND JUSTICE: A REVIEW OF RESEARCH, vol. 22 (M. Tonry ed., Chicago: University of Chicago Press, 1997).

27. These and other incidents involving racist speech are described, and attributed to various sources, in Charles R. Lawrence III, *If He Hollers Let Him Go: Regulating Racist Speech on Campus*, DUKE L. J. 431–83 (1990), reprinted in WORDS THAT WOUND,. (M. Matsuda et. al. eds., Boulder: Westview Press, 1993).

28. See Gregory Herek, *Documenting Prejudice Against Lesbians and Gay Men on Campus: The Yale Sexual Orientation Survey*, 25 J. HOMOSEXUALITY 15–30 (1993).

29. *See, e.g.*, Kevin Berrill, *Antigay Violence and Victimization in the United States: An Overview*, *in* HATE CRIMES: CONFRONTING VIOLENCE AGAINST LESBIANS AND GAY MEN (Newbury Park, CA: Sage Publications, 1992); Antony D'Augelli, *Lesbian and Gay Male Undergraduates' Experiences of Harassment and Fear on Campus*, 7 J. INTERPERSONAL VIOLENCE 383–95 (1992); Arnold Barnes and Paul Ephross, *The Impact of Hate Violence on Victims: Emotional and Behavioral Responses to Attacks*, 39 SOC. WORK 247–51 (1994); Herek, *Documenting Prejudice Against Lesbian and Gay Men on Campus*; Gregory Herek et al., *Hate Crime Victimization Among Lesbian, Gay, and Bisexual Adults: Prevalence, Psychological Correlates, and Methodological Issues*, 12 J. INTERPERSONAL VIOLENCE 195–215 (1997); and Janet Swim, Laurie Cohen, and Lauri Hyers, *Experiencing Everyday Prejudice and Discrimination*, *in* PREJUDICE: THE TARGET'S PERSPECTIVE, (J. Swim and C. Stangor eds., San Diego: Academic Press, 1998).

https://doi.org/10.1017/S1352325201072019 Published online by Cambridge University Press

Some campus hate speech regulations, adopted as the result of such incidents, were rather sweeping and have already been struck down in the courts as over-broad.[30] But the interesting question is not whether there are some, perhaps over-broad, ways of conceiving of hate speech and regulating it that would run afoul of Millian or constitutional principles, but rather whether there are plausible ways of conceiving of hate speech and regulating it that do not run afoul of Millian or constitutional principles. To focus our attention profitably, I will discuss the most sensible hate speech regulation of which I am aware.

The Stanford University Discriminatory Harassment Provision provides one useful characterization of hate speech.[31] It prohibits harassment of students on the basis of gender, race, color, handicap, religion, sexual orientation, or national and ethnic origin. Prohibited harassment "includes discriminatory intimidation by threats of violence and . . . personal vilification of students."

> Speech or other expression constitutes harassment by personal vilification if it: (a) is intended to insult or stigmatize an individual or a small number of individuals on the basis of their sex, race, color, handicap, religion, sexual orientation, or national and ethnic origin; and (b) is addressed directly to the individual or individuals whom it insults or stigmatizes; and (c) makes use of insulting or "fighting" words or non-verbal symbols.

The Stanford regulation understands hate speech quite narrowly—more narrowly than many regulations that target discriminatory speech. We might provisionally characterize discriminatory speech as speech that reflects group stereotypes and represents groups or their members as inferior by virtue of these stereotypes. But not all discriminatory speech is hate speech, at least as Stanford understands it. According to the Stanford provision, hate speech must employ insulting or fighting words or non-verbal symbols in face-to-face encounters with the intent to insult or stigmatize

30. In Doe v. University of Michigan, 721 F. Supp. 852 (E.D. Mich. 1989), a University of Michigan regulation that prohibited any behavior that "stigmatizes or victimizes" an individual on the basis of race, gender, etc., and that has the "reasonably foreseeable effect of interfering with an individual's academic efforts" or "creates an intimidating, hostile or demeaning environment for educational pursuits" was struck down in district court as over-broad. In a pamphlet that was distributed by the university in explanation of the new regulation, the following example was given of proscribed behavior: "A male student makes remarks in class like 'Women just aren't as good in this field as men' thus creating a hostile learning atmosphere for female classmates." *Cf.* UWM Post v. Board of Regents of the University of Wisconsin, 774 F. Supp. 1163 (E.D. Wis. 1991) (granting summary judgment to uphold a district court ruling that the university's anti-discrimination rule was over-broad and not restricted to fighting words) and Dambrot v. Central Michigan University, 839 F. Supp. 477 (E.D. Mich. 1993) (striking down the university's discriminatory harassment policy as over-broad and void for vagueness).

31. The Stanford provision is reproduced, among other places, in Thomas Grey, *Civil Rights vs. Civil Liberties: The Case of Discriminatory Verbal Harassment*, 8 SOC. PHIL. & POL'Y 106–107 (1991). As I explain in Section X, the Stanford provision has been struck down by a state court.

https://doi.org/10.1017/S1352325201072019 Published online by Cambridge University Press

on the basis of membership in certain groups. So, for example, if a sociology professor endorses, in print or in the classroom, genetic explanations of patterns of crime among blacks, this does not constitute prohibited harassment.[32] A heterosexual can tell a homosexual that she is not entitled to the same rights as heterosexuals without violating the regulation. And epithets (e.g., "nigger," "kike," and "faggot") that could not be employed by one student against another can be employed in a public forum, for instance, at a campus rally organized by neo-Nazis.[33]

Some aspects of the Stanford provision deserve comment or qualification. The provision does not expressly require that the speech be experienced as insulting or stigmatizing by the persons to whom it is addressed. However, the provision does require that the speech employ fighting words or non-verbal symbols, and fighting words, we have seen, are those epithets that "by their very utterance inflict injury or tend to incite an immediate breach of the peace." Presumably we should understand fighting words, as *Chaplinsky* does, as those that would have such immediate effects on a *reasonable person*.[34] Also, whether certain words or symbols would injure or provoke a reasonable person will depend upon various *contextual factors*. For example, fighting words will tend to be more injurious or provocative if directed at a member of a historically marginalized group by a member of a historically privileged group than otherwise or if unprovoked than if provoked. Further, insofar as it requires that speakers intend to insult or stigmatize, the provision may be too narrow. It ought to be enough that the speaker could *reasonably foresee* that his speech or symbols are by their nature such as to injure or provoke a reasonable person. And one might wonder whether the provision's restrictions to individual and small group encounters, contained in clauses (a) and (b), are compelling. It is not clear that hateful words are less injurious or provocative when directed at a group of which one is a member than when directed at oneself in particular. If so, one might contemplate a broader regulation covering public speech that is directed at large groups as well as speech directed at particular individuals. Finally, we might consider restricting hate speech to speech that is suitably hard to avoid and so involves a captive audience. Espe-

32. *Cf.* "Contrast the following two statements addressed to a black freshman at Stanford: (A) LeVon, if you find yourself struggling in your classes here, you should realize it isn't your fault. It's simply that you're the beneficiary of a disruptive policy of affirmative action that places underqualified, underprepared and often undertalented black students in demanding educational environments like this one. The policy's egalitarian aims may be well-intentioned, but given the fact that aptitude tests place African Americans almost a full standard deviation below the mean, even controlling for socioeconomic disparities, they are also profoundly misguided. The truth is, you probably don't belong here, and your college experience will be a long downhill slide. (B) Out of my face, jungle bunny." Henry Louis Gates, Jr. *Let Them Talk*, NEW REPUBLIC Sept. 20 and 27, 1993, at 45. The Stanford provision prohibits (B), not (A).

33. *See* Grey, *Civil Rights vs. Civil Liberties*, 94–95.

34. *See Chaplinsky*, at 573, endorsing the state court interpretation of the test of offensiveness in terms of the reaction of "men of common intelligence," rather than the reaction of the actual audience. Concern with the immediate effects on a reasonable person reduces worries that the provision gives the thin-skinned a veto on speech, akin to an unacceptable "heckler's veto" on speech.

https://doi.org/10.1017/S1352325201072019 Published online by Cambridge University Press

cially in public fora, it is common to be more concerned about the effects of offensive speech on a captive audience.[35] It is not, I think, that the magnitude of the offense or injury need be any smaller when inflicted on a non-captive audience, which could have avoided the speech but did not, than when inflicted on a captive audience. Rather, the offense or injury seems less morally significant if it could have been readily avoided. Although the sorts of incidents described above involve face-to-face encounters in which the victim presumably was a captive audience, face-to-face encounters need not involve a captive audience if the audience can readily anticipate and avoid the encounter. We might, therefore, consider restricting hate speech to speech, whether addressed to particular individuals or groups, that is addressed to a captive audience.[36] So whereas we might expand the provision to cover public speech, we might narrow it to require a captive audience.[37] It might be useful, for purposes of discussion, to construct a hate speech regulation that is modeled on the Stanford provision but attempts to incorporate these comments and qualifications. As a first approximation, such a provision might read as follows:

> Speech or other expression constitutes prohibited harassment by personal vilification if and only if: (a) it employs fighting words or non-verbal symbols that insult or stigmatize persons on the basis of their gender, race, color, handicap, religion, sexual orientation, or national and ethnic origin; (b) it is addressed to a captive audience; (c) the insult or stigma would be experienced by a reasonable person in those circumstances; and (d) it would be reasonable for the speaker to foresee that his words would have these effects on a reasonable person in those circumstances.

35. *Cf.* Lehman v. Shaker Heights, 418 U.S. 298 (1974) (upholding a city policy disallowing controversial political and public service messages in city buses as protecting captive audiences).

36. Another—more expansive—conception of hate speech would not make the captive audience requirement a condition of hate speech but would recognize that, all else being equal, hate speech is worse if directed at a captive audience.

37. Hate speech that does not involve a captive audience would not fall within the scope of such a regulation. How such a regulation would apply to the sort of speech at issue in Skokie is not entirely clear. *See* Skokie v. National Socialist Party of America, 373 N.E.2d 21 (1978); Collin v. Smith, 578 F.2d 1197 (7th Cir.) (1978); Collin v. Smith, 439 U.S. 916 (1978) (denying efforts to enjoin neo-Nazi groups from an organized demonstration in Skokie, Ill—a community with a large Jewish population, including many Holocaust survivors—as inconsistent with freedom of expression). The proposed march was to have taken place in a public forum (on public streets) and was advertised in advance, thus making it possible for offended parties to avoid the march or to counter with opposing speech. Insofar as these things are true, it might be argued that there was no captive audience and that the planned Skokie demonstration would not run afoul of a suitably restricted hate speech regulation. However, there is a very real sense in which citizens of Skokie were a captive audience. Given that the courts allowed the neo-Nazis to march in Skokie, it would be facile to suppose that Holocaust survivors in Skokie could avoid experiencing trauma simply by shutting their blinds. These captive audience aspects of the Skokie controversy argue for letting the neo-Nazis march, but elsewhere. *Cf.* Frisby v. Shultz, 487 U.S. 474 (upholding an ordinance prohibiting residential picketing aimed at a particular residence as a legitimate protection for captive audiences provided it leaves open ample alternative channels of communication).

https://doi.org/10.1017/S1352325201072019 Published online by Cambridge University Press

Because of its obvious debt to the Stanford provision, we might call this the neo-Stanford provision. It articulates a conception of prohibited speech but it does not itself specify sanctions for noncompliance. Reasonable sanctions would vary with the severity of the infraction but would presumably include reprimands, disqualification from certain campus honors and privileges, and other sanctions typically employed for infractions of academic conduct codes.

VI. MILLIAN PRINCIPLES AND HATE SPEECH

One can easily understand how Millian principles might make hate speech regulations seem suspect. Freedom of expression is among the most fundamental of our constitutional liberties because of the central role open discussion plays in the proper exercise of our capacities for responsible agency. This is why content-specific restrictions on speech tend to be especially problematic.[38] Of course, neither Millian nor constitutional principles imply that fundamental liberties can never be restricted. For Mill, the harm principle limits freedom of expression; if my speech is very likely to cause significant harm to others, then my speech may be restricted. As a constitutional matter, even high-value speech may be restricted provided the state does so in the pursuit of a compelling interest and in the least restrictive manner possible. So, for example, public demonstrations that posed a clear and present danger could be permissibly restricted, and Mill would agree (iii 1). But it may seem that hate speech regulations do not fall within the class of legitimate restrictions of fundamental liberties, for they seem to be designed to prevent offense. Mill, as we have seen, is quite explicit that mere offensiveness does not constitute harm for purposes of the harm principle; rather, conduct must harm important interests to count as harmful. And it would seem difficult to recognize a compelling state interest in preventing offense. Rights are considerations with a certain dialectical force in matters of political morality. They trump or constrain arguments of convenience or utility. If someone has a right to something, then it would be wrong to deprive her of it merely on the ground that doing so would be useful or would avoid offense. If constitutional guarantees are rights, then it cannot be a sufficient warrant to infringe them that doing so would avoid offense or discomfort to others. Discomfort and offense are the price we must pay for constitutional rights. Justice Brennan applies this principle to First Amendment liberties, in the process of overturning a flag desecration conviction in *Texas v. Johnson*, when he writes: "If there is a bedrock principle underlying the First Amendment, it is that the

38. For instance, Robert Post rests his skepticism about hate speech regulation on a commitment to deliberative values associated with democratic self-governance. *See* Robert Post, *Free Speech and Religious, Racial, and Sexual Harassment: Racist Speech, Democracy, and the First Amendment*, 32 WM. & MARY L. REV. 267–327 (1991).

https://doi.org/10.1017/S1352325201072019 Published online by Cambridge University Press

Government may not prohibit the expression of an idea simply because society finds the idea itself offensive or disagreeable."[39]

Mill does consider restrictions on "intemperate" speech that exceeds "the bounds of fair discussion" (ii 44). He observes that there is more to be said on behalf of such restrictions when they are applied to the expression of prevailing views than when they are applied to the expression of minority views:

> In general opinions contrary to those commonly received can only obtain a fair hearing by studied moderation of language and the most cautious avoidance of unnecessary offense, from which they hardly ever deviate even in a slightest degree without losing ground, while unmeasured vituperation employed on the side of the prevailing opinion really does deter people from professing contrary opinions and from listening to those who profess them. For the interest, therefore, of truth and justice it is far more important to restrain this employment of vituperative language than the other. . . . (ii 44)

But he ultimately rejects all such restrictions, claiming that it is "obvious that law and authority have no business . . . restraining either" (ii 44). This may make it seem clear that Mill is committed to rejecting hate speech regulation.

But notice that Mill conceives of intemperate speech as speech that employs "invective, sarcasm, personality and the like" (ii 44). But then the class of intemperate speech is much broader than the class of hate speech. So one might well reject sweeping regulation of intemperate speech, while accepting narrower regulation of hate speech. Moreover, though Mill rejects restrictions on intemperate speech, he does (in the passage just quoted) recognize special deliberative reasons for such restrictions when applied to the expression of prevailing views. His claim is that intemperate speech, especially directed against minority views, tends to stifle expression of minority views or prevent them from receiving a fair hearing. Whereas these deliberative interests in regulating intemperate speech may not be sufficient to justify such extensive restrictions, similar deliberative interests may be sufficient to justify narrower restrictions on hate speech. Indeed, Mill seems to recognize a "fighting words" exception to his usual prohibition on censorship:

> [E]ven opinions lose their immunity when the circumstances in which they are expressed are such as to constitute their expression a positive instigation to some mischievous act. An opinion that corn dealers are starvers of the poor, or that private property is robbery, ought to be unmolested when simply circulated through the press, but may justifiably incur punishment when delivered orally to an excited mob assembled before the house of a corn dealer, or when handed about among the same mob in the form of a placard. (iii 1)

39. 491 U.S. 397 (1989) (invalidating a statute prohibiting flag desecration statute as applied to someone who burned the American flag as an expression of political protest) at 414.

https://doi.org/10.1017/S1352325201072019 Published online by Cambridge University Press

To see how Mill might defend hate speech regulation despite his opposition to various forms of censorship, it is instructive to consider his views about paternalism. Recall that Mill's initial statement of the simple principle underlying his essay is that restrictions on liberty are permissible just in case they are applications of the harm principle; on this principle, paternalism is never justified. But Mill's statement of his own view is over-simple; he is forced to qualify his blanket prohibition on paternalism in order to maintain his claim that no one should be free to sell herself into slavery:

> The ground for thus limiting his power of voluntarily disposing of his own lot is apparent, and is very clearly seen in this extreme case. . . . [B]y selling himself for a slave, he abdicates his liberty; he foregoes any future use of it beyond that single act. He, therefore, defeats in his own case, the very purpose which is the justification of allowing him to dispose of himself. (v 11)

If we understand Mill's usual objections to paternalism in terms of his appeal to deliberative values, then we can see why he would make this exception to his usual anti-paternalism. Because it is the importance of exercising one's deliberative capacities that explains the importance of certain liberties, the usual reason for recognizing liberties provides an argument against extending liberties to do things that will permanently undermine one's future exercise of those same capacities. In this case, an exception to the usual prohibition on paternalism is motivated by appeal to the very same deliberative values that explain the usual prohibition. Perhaps a similar exception can be made in the case of hate speech regulation to the usual prohibition on censorship. Can hate speech regulation be shown to protect or advance the very deliberative values that explain why censorship is usually impermissible? If so, then this has an important bearing on the constitutional status of hate speech regulations as well. For not all content-specific restrictions on speech raise serious First Amendment concerns. The liberty to engage in low-value speech, for instance, involving libel or fighting words, is not a fundamental liberty, and, as a result, restrictions on low-value speech need not satisfy strict scrutiny or similarly demanding standards. The rationale for this treatment is that low-value speech does not engage deliberative values that underlie central First Amendment liberties. So if hate speech does not advance but retards deliberative values, then it is low-value speech, and hate speech regulation need not satisfy strict scrutiny or similarly demanding standards.

VII. IS HATE SPEECH HIGH-VALUE SPEECH?

In deciding whether the liberty to engage in hate speech is a fundamental one, it is important to bear in mind how narrowly the Stanford and neo-Stanford provisions conceive of hate speech. There is much speech that is discriminatory but does not count as hate speech. It reflects and encourages

https://doi.org/10.1017/S1352325201072019 Published online by Cambridge University Press

bias and harmful stereotyping, but it does not employ epithets in order to stigmatize and insult. Though pernicious, this broader class of discriminatory speech would often be part of an open debate about matters of political access and opportunity, distributive justice, and other matters of public and private concern. For instance, it is difficult to see how one could have an open and vigorous discussion of the merits of affirmative action without allowing the expression of views that might reflect and encourage racial stereotyping. By contrast, it does not seem difficult to imagine an open and vigorous discussion of affirmative action that did not permit the use of racial epithets to vilify and wound. Moreover, it seems plausible that, all else being equal, hate speech is worse than discriminatory speech. In contrast with merely discriminatory speech, hate speech's use of traditional epithets or symbols of derision to vilify on the basis of group membership expresses contempt for its targets and seems more likely to cause emotional distress and to provoke visceral, rather than articulate, response. Whereas it is difficult to treat all discriminatory speech as low-value,[40] hate speech can plausibly be claimed to retard rather than advance deliberative values.[41]

Consider the immediate effects of hate speech on deliberative interests. Hate speech, like speech employing (other) fighting words, expresses its speaker's hostility and disrespect for its intended audience but does not articulate the grounds for the speaker's perspective and attitudes or a proposal for debate and decision. Moreover, also like fighting words, hate speech evokes visceral, rather than articulate, response; it provokes violence or, more commonly, silences through insult or intimidation. Charles Lawrence effectively describes the visceral and inarticulate reaction hate speech often evokes:

> . . . visceral emotional response to personal attack precludes speech. Attack produces an instinctive, defensive psychological reaction. Fear, rage, shock, and flight all interfere with any reasoned response. Words like "nigger," "kike," and "faggot" produce physical symptoms that temporarily disable the victim, and the perpetrators often use these words with the intention of producing this effect. Many victims do not find words of response until well after the assault, when the cowardly assaulter has departed.[42]

This explains well why it is perhaps psychologically somewhat naïve to suppose that in this case the corrective to bad speech is more speech; hate

40. This is why the very sweeping prohibition on discriminatory speech contained in Article 4 of the U.N. Convention on the Elimination of All Forms of Racial Discrimination seems over-broad. It is significant that some countries, such as Great Britain and Germany, that have sought to comply in some way with Article 4 have rejected the Article 4 prohibition as too broad and have instead endorsed narrower regulations targeting some form of hate speech. *See Positive Measures Designed to Eradicate Racial Discrimination*, at 23–25.

41. In Section X, I discuss the related question, raised in R.A.V. v. St. Paul, 505 U.S. 377 (1992), of whether and, if so, under what conditions it is permissible to regulate a subclass of speech (e.g., hate speech) that is part of a broader class of problematic speech (e.g., discriminatory speech or fighting words).

42. Lawrence, *If He Hollers Let Him Go*, at 68.

https://doi.org/10.1017/S1352325201072019 Published online by Cambridge University Press

speech typically preempts further speech. When discussion turns to name-calling, reasoned debate becomes difficult, if not impossible. Provided that we understand fighting words, as *Chaplinsky* does, to include those words that silence through injury or intimidation, as well as those that provoke fights, hate speech, understood in something like the way the Stanford or neo-Stanford provision does, would seem to be a case of fighting words.[43] Moreover, because hate speech takes place against a background of historical and ongoing discrimination and violence, it seems more likely than other forms of fighting words to elicit visceral reactions of rage, fear, or intimidation and silence.[44] Insofar as hate speech, like fighting words, expresses visceral attitudes and elicits inarticulate reactions, it doesn't engage deliberative values central to Millian and constitutional principles that normally protect speech.

Arguably, there are also other negative but more remote effects of hate speech on deliberative interests. As Mill argues, effective deliberation requires reflective consideration and assessment of alternatives, and this requires an open-minded discussion of diverse viewpoints. As his remarks on intemperate speech (ii 44) suggest, it is arguable that the proper representation and evaluation of diverse viewpoints require a background culture of mutual respect among members of the deliberative community. But hate speech can poison the well of mutual respect and discourage participation in the deliberative community. There are a couple of ways in which the divisiveness of hate speech threatens the development of a deliberative culture. First, a robust deliberative culture requires widespread participation, especially by people with diverse viewpoints. But there is some reason to think hate speech often not only preempts immediate reply but also discourages targets from participation in public deliberations. Insofar as hate speech tells some members of the community of inquirers that they do not belong or are not members in good standing, it is likely to discourage their full deliberative engagement. Indeed, there is significant empirical literature showing that clear discriminatory treatment often leads targets of discrimination to withdraw from or avoid contexts of potential discrimina-

43. Nadine Strossen complains that whereas hate speech regulations are concerned to prevent emotional injury, the fighting words doctrine is concerned with violence speech may cause; *see* Nadine Strossen, *Regulating Racist Speech on College Campus: A Modest Proposal?* DUKE L. J. 515 (1990). But whereas this claim about the scope of the fighting words doctrine might be plausible on narrowly doctrinal grounds, it has no strong principled rationale. As I argue in Section IV, *Chaplinsky* rightly construes fighting words more broadly; if failure to engage deliberative values is the reason fighting words are low-value, then there is as much reason to treat words that silence by psychic injury as low-value as there is to treat words that provoke fisticuffs as low-value.

44. *See, e.g.*, Kent Greenwalt, *Reflections on Justifications for Defining Crimes by the Category of Victim*, ANN. SURV. AM. L. 617–28 (1992–93); Herek et al., *Hate Crime Victimization Among Lesbian, Gay, and Bisexual Adults* (finding that the psychological consequences of victimization are greater for targets of bias crimes than they are for the targets of comparable non-bias crimes); and Craig, *Retaliation, Fear, or Rage* (finding that African-American targets were more likely to express a desire for revenge in bias crimes than in comparable non-bias crimes).

https://doi.org/10.1017/S1352325201072019 Published online by Cambridge University Press

tion.[45] Insofar as targets of hate speech withdraw from participation in deliberative contexts, this adversely affects their deliberative interests and the quality of the deliberations that take place without their input. Second, as Mill's remarks about intemperate speech suggest, hate speech may well prevent targets from receiving a fair hearing even when they do participate. An atmosphere charged with expression of contempt is not conducive to a careful and impartial evaluation of the target's contributions, and there is empirical literature suggesting that overheard slurs adversely affect how audiences evaluate the contributions of targets.[46]

In these ways, hate speech contributes to a hostile environment that undermines the culture of mutual respect necessary for effective expression and fair consideration of diverse points of view. This is one sense in which hate speech silences its victims.[47] If so, this is a deliberative cost that all members of the community, including hate speakers themselves, and not

45. *See, e.g.*, Barnes and Ephross, *The Impact of Hate Violence on Victims*; Herek et al, *Hate Crime Victimization Among Lesbian, Gay, and Bisexual Adults*; Swim, Cohen, and Hyers, *Experiencing Everyday Prejudice and Discrimination*. Several studies suggest that discriminatory attitudes within a domain often lead targets to disengage from that domain as a means of preserving self-esteem. *See, e.g.*, Claude Steele, *A Threat in the Air: How Stereotypes Shape Intellectual Identity and Performance*, 52 AM. PSYCHOLOGIST 613–29 (1997); Claude Steele and Joshua Aronson, *Stereotype Threat and the Intellectual Test Performance of African Americans*, 69 J, PERSONALITY & SOC. PSYCHOLOGY 797–811 (1995); Joshua Aronson, Diane Quinn, and Stephen Spencer, *Stereotype Threat and the Academic Underperformance of Minorities and Women*, in PREJUDICE: THE TARGET'S PERSPECTIVE; Kevin Allison, *Stress and Oppressed Social Category Membership*, in PREJUDICE: THE TARGET'S PERSPECTIVE; Brenda Major and Toni Schmader, *Coping with Stigma Through Psychological Disengagement*, in PREJUDICE: THE Target's PERSPECTIVE; and Nyla Branscombe and Naomi Ellemers, *Coping with Group-Based Discrimination: Individualistic Versus Group-Based Strategies*, in PREJUDICE: THE Target's PERSPECTIVE.

46. *See* Jeff Greenburg and Tom Pyszczynski, *The Effect of an Overheard Ethnic Slur on Evaluations of Target: How to Spread a Social Disease*, 21 J. EXPERIMENTAL SOC. PSYCHOLOGY 61–72 (1985), and Jeff Greenburg, S.L. Kirkland, and Tom Pyszczynski, *Some Theoretical Notions and Preliminary Research Concerning Derogatory Ethnic Labels, in* DISCOURSE AND DISCRIMINATION, (G. Smitherman-Donaldson and T. van Dijk eds., Detroit: Wayne State University Press, 1988).

47. Catherine MacKinnon and others have suggested that pornography silences women because it creates conditions that make it impossible for women to perform certain kinds of speech acts (e.g. to perform a speech act signifying refusal to consent to sex) and that, as a result, women have First Amendment claims against pornographers. *See* CATHERINE MacKINNON, ONLY WORDS (Cambridge: Harvard University Press, 1993); *cf.* Rae Langton, *Speech Acts and Unspeakable Acts*, 22 PHIL. & PUB. AFF. 293–330 (1993). In claiming that hate speech can and does silence its victims, I am not claiming that hate speech makes counter-speech by its victims impossible or that it gives the victims of hate speech a First Amendment claim against hate speakers. These claims about silencing do not strike me as very plausible as applied either to pornography or to hate speech. Rather, my claim is the Millian one that hate speech, especially when directed by a member of a dominant group against a member of a historically marginalized group, tends to provoke immediate responses that are inarticulate and tends to stifle the expression of marginalized points of view or prevent them from receiving a fair hearing. If this is so, it's not that the victims of hate speech have a First Amendment claim against hate speakers (a claim that would seem to run afoul of the familiar observation that the First Amendment protects the speech of individuals against state, rather than private, action). It's rather that hate speech retards the very deliberative values that explain why content-specific censorship is normally impermissible; if this is the case, then hate speech regulation is a well-motivated exception to the usual prohibition on content-specific censorship.

https://doi.org/10.1017/S1352325201072019 Published online by Cambridge University Press

just their targets and others, must pay for unregulated hate speech.[48] If the value of liberties of expression depends upon a culture of mutual respect, then considerations of equality—the requirements of equal concern and respect—constrain the kinds of liberties that can plausibly be regarded as fundamental.[49]

If hate speech retards deliberative values, and hate speech regulation protects deliberative values, then we should not see hate speech regulations like the Stanford and neo-Stanford provisions as restricting fundamental liberties. Hate speech regulation can be seen as a well-motivated exception to the usual prohibition on content-specific regulation of speech.

However, it might be claimed that, even if many of the immediate and remote effects of hate speech do not engage and even frustrate deliberative values, some of its remote effects do or at least can contribute to deliberative values. Hate speech can and does provoke reasoned debate typically by others at a later time about the importance and grounds of tolerance and mutual respect and the permissible ways of promoting these values. For instance, the public debate about hate speech and its regulation is itself a response to hate speech.

At most, this shows that the effects of hate speech on deliberative values are mixed; whereas hate speech marginalizes its victims and other group members within the deliberative community, it can occasion reasoned debate about tolerance and mutual respect at a later time. However, there's something peculiar about claiming that hate speech has or can have this sort of deliberative value of producing correctives to hate speech. It seems a bit like valuing poverty as a condition that makes charity possible. In both cases, it seems that the underlying values would be better served if the condition requiring correction never occurred in the first place. Moreover, there is an important kind of correction that further speech—condemning hate speech—seems ill suited to effect. Scholarly or even public debate that is occasioned by hate speech cannot prevent the emotional distress that hate speech tends to cause to targets, nor is it likely to prevent targets from feeling marginalized within the deliberative community. These considerations lead me to doubt whether the deliberative value of hate speech is genuine or, if genuine, commensurate with its deliberative costs. If the overall effect of hate speech on deliberative values is negative, we should not recognize hate speech as high-value speech.

Does this mean that hate speech is low-value speech? That depends. First Amendment jurisprudence is generally *bivalent* in the sense that high-value

48. The claim that hate speakers, as well as hate listeners, pay a deliberative cost for hate speech is like Mill's claim that exploiters and oppressors are themselves harmed by exploitative relationships, because such relationships are inconsistent with beneficial personal and social interaction (*SW:* iv 522–5, 541).

49. Whereas I do mean to suggest that basic liberties are conditioned by, and so posterior to, a conception of equality as involving equal concern and respect for deliberative agents, I do not mean to suggest that basic liberties are conditioned by a substantive egalitarian conception of distributive justice. That would require considerably more argument.

https://doi.org/10.1017/S1352325201072019 Published online by Cambridge University Press

and low-value speech are treated as mutually exclusive and jointly exhaustive categories. On this view, if hate speech is not high-value, it must be low-value. But bivalence is not a necessary feature of First Amendment jurisprudence. An alternative approach is *scalar;* one assesses liberties of expression in terms of their *centrality* to First Amendment values and holds regulation of expression to a standard of review whose stringency is commensurate with the importance of the liberties at stake.[50] Indeed, First Amendment doctrine is not consistently bivalent; under existing law, commercial speech such as advertising is treated as neither high-value nor low-value; it is accorded a kind of intermediate value, with the result that restrictions on commercial speech must satisfy an intermediate standard of review.[51] Even on this scalar view, a good case can be made that hate speech is low-value speech. But perhaps this claim is more controversial on the scalar than on the bivalent view. However, even if we do not agree to treat hate speech as low-value speech, the case for doing so makes it very implausible to treat it as high-value speech. At most, it could be accorded a sort of intermediate value; if so, the regulation of hate speech should be subject, at most, to a standard of review intermediate between rational basis review and strict scrutiny—perhaps one that requires that the state have a significant or substantial interest that it pursues in a comparatively restrictive manner.

VIII. CAN HATE SPEECH REGULATIONS SATISFY THE RELEVANT STANDARD OF REVIEW?

We must determine not only whether hate speech is high-value or low-value speech but also whether hate speech regulations can satisfy the associated standard of review. Of course, if, as I have argued, hate speech is not high-value speech, then that makes its regulation considerably easier to justify. However, even if we did recognize hate speech as high-value speech, a good case could still be made for the permissibility of hate speech regulation.

Mill thinks that even fundamental liberties can be restricted when their

50. The scalar approach is defended by Justice Stevens in his concurring opinion in *R.A.V.*

51. Virginia State Board of Pharmacy v. Virginia Citizens Consumer Council, 425 U.S. 748 (1976) (recognizing commercial speech as protected by the First Amendment and invalidating a Virginia statute prohibiting pharmacists from advertising prescription drug prices); Central Hudson Gas v. Public Service Commission of New York, 447 U.S. 557 (1980) (articulating the following four-part test for commercial speech—(1) the speech must concern lawful activity and not be misleading, (2) there must be a substantial governmental interest in its regulation, (3) the regulation must directly advance the governmental interest, and (4) there must not be an equally good but less restrictive means for achieving the governmental interest—and invalidating a prohibition on advertising within the electrical utility industry designed to increase electricity consumption as not satisfying the least restrictive means test [(iv)]); and Board of Trustees of SUNY v. Fox, 492 U.S. 469 (1989) (interpreting *Hudson*'s least-restrictive-means test as a substantial-fit test and upholding a university prohibition on the operation of commercial enterprises within student dormitories).

https://doi.org/10.1017/S1352325201072019 Published online by Cambridge University Press

exercise would cause harm to others. As we have already noted, he is quite explicit that mere offensiveness does not constitute harm for purposes of the harm principle; rather, conduct must adversely affect important interests to count as harmful. I might be offended by your religious or sexual practices, but that does not show that your practices harm me (*cf. OL:* iv 14–16, 20–21). That seems right. But it would be a mistake to suppose that the effect of hate speech is mere offense.[52]

Mere offense typically involves non-traumatic psychological disturbance that is both mild and ephemeral and does not ramify strongly to other psychological states or to behavior. But the effects of hate speech on targets are often traumatic. For one thing, one would expect targets of hate speech to experience many of the same sort of short-term and long-term psychological effects common to victims of various kinds of crimes. An American Psychological Association Task Force on the Victims of Crime and Violence found that the victims of diverse personal crimes had quite similar reactions, including anger, fear, anxiety, helplessness, sleep deprivation, depression, loss of self-esteem, and a deterioration of personal relationships.[53] Research suggests that bias-motivated harassment does commonly produce just such short-term and long-term effects.[54] Moreover, because hate speech takes place against a background of historical and ongoing discrimination and violence, there is reason to expect that it is more likely than other forms of verbal harassment to produce such effects.[55] In particular, there is considerable evidence that bias-motivated harassment leads targets to modify their aims and behavior—for instance, where they want to go, what they want to do, with whom they want to associate, where they seek success, and how they evaluate themselves—in various ways so as to avoid potential discrimination.[56] Moreover, bias-motivated harassment tends to reduce the self-esteem of targets; in order to avoid adverse effects on self-esteem, targets tend, again, to modify their

52. In several places, Ronald Dworkin appeals to the principle that offense and disgust are insufficient grounds for restricting speech in order to condemn hate speech regulation and the regulation of pornography. *See* RONALD DWORKIN, FREEDOM'S LAW 204–205, 218–19, and 237–38 (Cambridge: Harvard University Press, 1996). If Dworkin thinks this principle undermines narrowly crafted hate speech regulations, such as the Stanford provision, as some of his remarks suggest, then I think he mistakes harm for offense. However, at one point, Dworkin does suggest that a narrowly crafted hate speech regulation, along the lines of the Stanford provision, would regulate speech outside the scope of protected freedoms; *see* Freedom's LAW, at 255–56.

53. AMERICAN PSYCHOLOGICAL ASSOCIATION, FINAL REPORT OF THE TASK FORCE ON THE VICTIMS OF CRIME AND VIOLENCE 25 (Washington, D.C.: American Psychological Association, 1984). *See also* Jacobs and Potter, *Hate Crimes*, at 31.

54. *See, e.g.*, Barnes and Ephross, *The Impact of Hate Violence on Victims*, and Herek et al., *Hate Crime Victimization Among Lesbian, Gay, and Bisexual Adults*.

55. *See, e.g.*, Greenwalt, *Reflections on Justifications for Defining Crimes by the Category of Victim*, and Herek et al., *Hate Crime Victimization Among Lesbian, Gay, and Bisexual Adults*; contrast Barnes and Ephross, The Impact of Hate Violence on Victims.

56. *See, e.g.*, AMERICAN PSYCHOLOGICAL ASSOCIATION, TASK FORCE ON THE VICTIMS OF CRIME AND VIOLENCE, at 25–27; Barnes and Ephross, *The Impact of Hate Violence on Victims*; and Swim, Cohen, and Hyers, *Experiencing Everyday Prejudice and Discrimination*."

https://doi.org/10.1017/S1352325201072019 Published online by Cambridge University Press

aims and behavior in various ways.[57] These findings suggest that the psychic costs that targets of hate speech bear often are significant and nonephemeral and have important psychological ramifications. These psychic costs go beyond mere offense; they are no less harmful, for being psychic.[58]

Indeed, these costs arguably can constitute emotional distress that is actionable in tort law.[59] According to the *Restatement of Torts:*

> Complete emotional tranquility is seldom attainable in this world, and some degree of transient and trivial emotional distress is part of the price of living among people. The law intervenes only where the distress is so severe that no reasonable man could be expected to endure it.[60]

Recovery is possible for the tort of infliction of severe emotional distress, provided the effects are foreseeable and result from behavior that is "so outrageous in character, and so extreme in degree, as to go beyond all possible bounds of decency, and to be regarded as atrocious, and utterly intolerable in a civilized community."[61] It does not seem unreasonable to regard the sort of hate speech incidents described above as intolerable in a civilized society regulated by principles of mutual tolerance and respect or

57. On adverse consequences for self-esteem, *see, e.g.*, Hope Landrine and Elizabeth Klonoff, *The Schedule of Racist Events: A Measure of Racial Discrimination and a Study of Its Negative Physical and Mental Health Consequences*, 22 J. BLACK PSYCHOLOGY 144–68 (1996); Hope Landrine et al., *Physical and Psychiatric Correlates of Gender Discrimination: An Application of the Schedule of Sexist Events*, 19 PSYCHOLOGY WOMEN Q. 473–92 (1995); and Herek et al, *Hate Crime Victimization Among Lesbian, Gay, and Bisexual Adults*. On disengagement as a way of protecting self-esteem, *see, e.g.*, Steele, *A Threat in the Air: How Stereotypes Shape Intellectual Identity and Performance*; Steele and Aronson, *Stereotype Threat and the Intellectual Test Performance of African Americans*; Aronson, Quinn, and Spencer, *Stereotype Threat and the Academic Underperformance of Minorities and Women*; Allison, *Stress and Oppressed Social Category Membership*; Major and Schmader, *Coping with Stigma Through Psychological Disengagement*; and Branscombe and Ellemers, *Coping with Group-Based Discrimination*. Some (*e.g.*, Barnes and Ephross, *The Impact of Hate Violence on Victims* and Craig, *Retaliation, Fear, or Rage*) deny finding greater adverse effects on self-esteem for targets in cases of bias-motivated crime, but this finding is compatible with the disengagement effect.

58. *See also* Frederick Schauer, *The Phenomenology of Speech and Harm*, 103 ETHICS 635–53 (1993), and Susan Brison, *Speech, Harm, and the Mind-Body Problem in First Amendment Jurisprudence*, 4 LEGAL THEORY 39–61 (1998). Contrast Larry Alexander, *Banning Hate Speech and the Sticks and Stones Defense*, 13 CONST. COMMENTARY 71–100 (1996).

59. *Cf.* Richard Delgado, *Words that Wound: A Tort Action for Racial Insults, Epithets, and Name Calling*, reprinted in WORDS THAT WOUND, (Matsuda et al. eds.); and Michael Steenson, *Civil Actions for Emotional Distress and R.A.V. v. City of St. Paul*, 18 WM. MITCHELL L. REV. 983–99 (1992). Hustler Magazine v. Falwell, 485 U.S. 46 (1988) (preventing recovery by a public figure for the tort of intentional infliction of emotional distress for publications of the sort at issue without showing that the publication contained false statements of fact that were made with knowledge that the statements were false or in reckless disregard of their truth) is compatible with this argument. *Hustler's* focus was on public figures, whereas there is no such focus in the hate speech provisions. Moreover, the hate speech provisions in question involve an unprotected class of speech, namely, fighting words; it is not clear that the parody at issue in *Hustler* involved fighting words.

60. RESTATEMENT (SECOND) OF TORTS § 46, comment (j) (1977).

61. RESTATEMENT § 46, comment (d).

https://doi.org/10.1017/S1352325201072019 Published online by Cambridge University Press

to regard the emotional injury that it causes as such that no reasonable person should be expected to endure.

When calculating the harm hate speech causes, we must reckon not only these psychic harms but also the adverse effects on deliberative interests that we already considered in arguing that hate speech is not high-value speech. Hate speech expresses visceral attitudes and invites inarticulate reactions, and it undermines the culture of mutual respect necessary for effective expression and fair consideration of diverse points of view. This is a deliberative cost that all members of the community pay, but the victims of hate speech clearly bear the biggest share of this cost. Because of the importance of deliberative interests in Mill's account of human happiness and in specifying fundamental interests and liberties, the adverse effects of hate speech on the deliberative interests of targets ought to be reckoned as harms.

These are good reasons for regarding hate speech as harmful and not merely offensive. But then the laissez-faire approach that leaves hate speech unregulated is costly. Moreover, it is worth noting that the costs of laissez-faire are not distributed equally.[62] Because hate speech is typically practiced against members of historically marginalized groups by people who are not members of such groups, and because in these cases the history of marginalization makes hate speech especially harmful, the costs of laissez-faire fall disproportionately on those who are already objects of discrimination and marginalization. Nor, I think, is there much reason to suppose that the unequal impact of laissez-faire will be equally distributed within the community over the long run. If the importance of First Amendment liberties implies that we must be prepared to pay a price for them, then the costs ought to be distributed equally or fairly. Insofar as the costs of the laissez-faire interpretation of First Amendment liberties are borne disproportionately by already marginalized groups, this casts serious doubt on that interpretation of fundamental First Amendment liberties.

These considerations about the kinds of harm that hate speech causes and the inequitable way in which they are distributed explain why a good case can be made for thinking that even if hate speech were high-value speech, narrowly crafted hate speech regulations could satisfy strict scrutiny. For it is reasonable to suppose that there is a compelling interest in preventing some citizens from inflicting emotional distress on others, in maintaining a climate of mutual tolerance and respect in which each can deliberate effectively about both private and public matters, and in ensuring a fair distribution of the costs of maintaining a system of fundamental liberties. Any hate speech regulation would have to satisfy the least restrictive means

62. *See* the useful discussion in Frederick Schauer, *Uncoupling Free Speech*, 92 COLUM. L. REV. 1321–57 (1992).

https://doi.org/10.1017/S1352325201072019 Published online by Cambridge University Press

test, but something very much like the Stanford or neo-Stanford provision would seem to meet this constraint.[63]

IX. PRAGMATIC CONCERNS

Even if principled objections to hate speech regulation can be met, there remains a common pragmatic worry. Some who oppose hate speech regulation do not claim that hate speech is high-value speech; rather, they claim that it is impossible to draw a clear line between high-value and low-value speech, with the result that one cannot regulate hate speech without jeopardizing or chilling genuinely high-value speech. Indeed, it is sometimes suggested that if we allow hate speech to be regulated we are on our way down a slippery slope toward regulating the speech rights of the very groups (e.g., blacks, gays, women) that are the targets of hate speech.[64] For example, it is claimed, protest by these marginalized groups and intra-group use of epithets—for instance, the use of the word "nigger" among blacks or the word "queer" among gays—would cease to be protected.

As a general matter, such pragmatic arguments are not especially compelling. Though we sometimes seek "bright-line" principles or categories in adjudication, we are always using the law to draw important distinctions between classes of conduct, even if this means there will be some difficult, borderline cases.[65] Imagine someone thinking that tort law should not allow recovery for negligence on the ground that once you allow someone to recover for reckless endangerment you are on your way down the slippery slope toward strict liability. Courts have articulated principles of negligence that allow them to do a reasonably good job of distinguishing between negligent and non-negligent infliction of harm and among various forms of negligence, even if some cases are difficult to decide. Nor does the

63. The hate speech provisions we have discussed target all hate speech based on gender, race, color, handicap, religion, sexual orientation, or national and ethnic origin—not just hate speech directed at a member of a historically marginalized group by a member of a historically advantaged group. This kind of viewpoint neutrality is reasonable insofar as the reasons for treating hate speech as low-value and for thinking that it causes harms that there is a compelling interest in preventing apply, at least in principle, to speech by members of historically marginalized groups as well. But, all else being equal, the case for regulating hate speech has special force when the speech is directed at a member of a historically marginalized group by a member of a historically advantaged group. For in these contexts, the history of marginalization, often by violence, affects the significance of epithets and non-verbal symbols; the tendency of hate speech in these circumstances to provoke visceral rather than articulate reaction, to wound, to threaten, and to poison the culture of mutual toleration and respect is greater, and these contextual factors should affect what sorts of speech a reasonable person would experience as insulting or stigmatizing. So even if hate speech provisions should be formulated in suitably viewpoint-neutral terms, their application should be guided by a robust sense of the significance of historical patterns of discrimination.

64. *Cf.* Strossen, *Regulating Racist Speech on College Campus*, at 512–13, 515, 526–30, 536–37, 556–58.

65. *Cf.* Irwin v. Gavit, 268 U.S. 161 (1925) at 168; and Frederick Schauer, *Slippery Slopes*, 99 HARV. L. REV. 361–83 (1985).

https://doi.org/10.1017/S1352325201072019 Published online by Cambridge University Press

pragmatic argument seem better in this particular area. We can distinguish between low-value and high-value speech in terms of deliberative values, and this allows us to explain why the Court does and should treat certain forms of speech as low-value. But the same principles suggest that hate speech, as conceived here, is low-value speech. Traditional forms of civic protest by marginalized groups will not count as hate speech. Such protest articulates and expresses moral claims against the community and so is a proper part of public deliberation about community action; it is typically publicly organized and does not involve a captive audience; and it typically does not employ epithets that aim to vilify its audience solely on the basis of group membership. Moreover, in the case of intra-group use of epithets, it is clear that there is no intention to insult or vilify on the basis of group membership and that it would not be reasonable for those to whom the speech is addressed to regard this use of epithets as insulting or provocative, and so this speech will not count as hate speech.

Indeed, there is a kind of pragmatic justification of hate speech regulation, understood as it has been here. The focus, within the class of discriminatory speech, on fighting words can be understood, in part, as selecting a bright-line principle. As we have seen, it is plausible to suppose that, all else being equal, hate speech is worse than discriminatory speech that does not employ fighting words, in the sense that the former tends more than the latter to cause those effects that make speech low-value and harmful. All else being equal, hate speech is more likely to cause emotional distress, to evoke visceral, rather than articulate response, and to stifle the expression of marginalized points of view or prevent them from receiving a fair hearing. But all else is not always equal; some instances of merely discriminatory speech surely have effects at least as bad as some instances of hate speech. If so, why target only hate speech? A partly pragmatic answer is that doing so focuses on a subcategory of discriminatory speech that is generally worse than other kinds of discriminatory speech and that is comparatively easy to identify; it may often be more problematic to identify other kinds of discriminatory speech as discriminatory or to ascertain the extent to which it produces the effects that make hate speech low-value and harmful. Focusing on hate speech gives us a comparatively bright line.

I am not suggesting that there won't be problems articulating and applying a plausible hate speech regulation. But I see no reason to think that the problems here will be any worse than those we already face in First Amendment jurisprudence and elsewhere. It must certainly be an empirical claim that it is harder to draw principled and salient lines here than elsewhere. Absent empirical support for this claim, it seems irresponsible to give up regulating hate speech without seriously trying. If it should turn out that our best efforts at a principled and salient conception of hate speech and its regulation not only have an unavoidably chilling effect on high-value speech but actually harm the legitimate speech interests of the very groups hate speech regulations are designed to protect, then that would be a good

https://doi.org/10.1017/S1352325201072019 Published online by Cambridge University Press

pragmatic objection to hate speech regulation. Hate speech regulation would then be a failed experiment. But the experiment has not even begun yet; claims that it fails are, therefore, premature.

X. THE STANFORD PROVISION AND R.A.V. v. ST. PAUL

Before concluding, I would like to discuss the Supreme Court's ruling in R.A.V. v. St. Paul and its bearing on hate speech regulation of the sort contained in the Stanford provision. In 1992 California enacted a statute (the so-called "Leonard law") that prohibits private universities from disciplining students for speech for which public universities could not discipline students.[66] A state court relied on *R.A.V.* to conclude that the Stanford provision would violate the First Amendment if Stanford were a public university, and therefore concluded that the Stanford provision violates the statute.[67] Unfortunately, Stanford did not appeal this decision.[68]

In 1990 the city of St. Paul, Minnesota adopted the Bias-Motivated Crime Ordinance that provides as follows:

> Whoever places on public or private property a symbol, object, appellation, characterization or graffiti, including, but not limited to, a burning cross or Nazi swastika, which one knows or has reasonable grounds to know arouses anger, alarm, or resentment in others on the basis of race, color, creed, religion or gender commits disorderly conduct and shall be guilty of a misdemeanor.[69]

R.A.V. was a juvenile who was convicted under the ordinance for burning a cross on the lawn of a black family. The Minnesota Supreme Court upheld the ordinance as constitutional, claiming that the phrase "arouses anger, alarm, or resentment in others" should be construed as restricting the reach of the ordinance to fighting words, within the meaning of *Chaplinsky*, and that the ordinance satisfied strict scrutiny.[70] The U.S. Supreme Court reversed. Justice Scalia delivered the opinion of the Court (in which Justices Rehnquist, Kennedy, Souter, and Thomas joined). Justices White, Blackmun, O'Connor, and Stevens concurred in the result but rejected Scalia's reasoning.

Scalia accepted, at least for the sake of argument, the Minnesota Supreme Court's interpretation of the ordinance as restricting only fighting words. He argued that even if this were so and one assumed fighting words were low-value speech, it wouldn't follow that the ordinance could regulate

66. Cal. Educ. Code § 94367.
67. Corry v. Stanford University, No. 740309 (Cal. Sup., Santa Clara Cty., Feb., 27, 1995).
68. For a useful discussion of the background to the Stanford provision and criticism of the state court decision, *see* Thomas Grey, *How to Write a Speech Code Without Really Trying: Reflections on the Stanford Experience*, 29 U.C. DAVIS L. REV. 891–956 (1996).
69. Minn. Legis. Code (1990) § 292.02.
70. In re Welfare of R.A.V., 464 N.W. 2d 507 (Minn. 1991).

https://doi.org/10.1017/S1352325201072019 Published online by Cambridge University Press

biased or hateful fighting words. Scalia's thesis is that even if the entire category of fighting words is unprotected speech, one cannot restrict only a proper part of this category on the basis of its content (fighting words that insult or provoke violence on the basis of race, color, creed, religion, or gender) unless one can show that the very reasons that make the class unprotected in the first place have special application to the proper subclass. Scalia assumed that no such rationale could be given and concluded that the ordinance's regulation of hateful fighting words must be subject to the same standard of review to which the regulation of protected speech is subject. He then went on to reject the Minnesota Supreme Court's claim that the ordinance satisfies strict scrutiny. Conceding that St. Paul has a compelling interest in preventing harassment of members of historically persecuted groups, Scalia claimed that it did not adopt the least restrictive means to achieve this end, because it could have adopted a topic-neutral restriction on fighting words.

I sympathize with those who concurred in the result but did not join in Scalia's opinion. They agreed that the ordinance is unconstitutional, but because it is over-broad. The ordinance prohibits any conduct that it is reasonable to expect would arouse anger, alarm, or resentment in others on the basis of race, color, creed, religion, or gender, and not just fighting words with this effect. As such, it targets discriminatory speech, not just fighting words or hate speech, and so restricts more than low-value speech.

But Scalia's thesis that content-specific restrictions within an unprotected class of speech are suspect is not compelling. We might agree that the state could adopt the more sweeping content-neutral restriction on fighting words. However, the question is whether there is anything impermissible in the state focusing on the salient subclass of hateful fighting words. As both White and Stevens observe, there are countless cases in which the Court has upheld selective content-specific restrictions within a class of non-fundamental liberties. We can illustrate this point outside First Amendment jurisprudence. In the post-*Lochner* era, liberty of contract is not regarded as a fundamental liberty, and so various forms of economic legislation designed to protect the interests of workers and to regulate industry standards are subject only to rational basis review.[71] Regulations targeted only at certain industries—for instance, the milk industry—are not thereby suspect; they need only pursue a legitimate state interest in a reasonable manner.[72] Within First Amendment jurisprudence, there are also various examples of legitimate sub-classification. As we have seen, commercial speech, such as advertising, is treated as neither high-value nor low-value; it is accorded a kind of intermediate value, with the result that restrictions on

71. Lochner v. New York, 198 U.S. 45 (1905) (invalidating a New York statute regulating the maximum number of hours that bakery employees could work as a violation of liberty of contract protected by the due process clause of the Fourteenth Amendment).

72. Nebbia v. New York, 291 U.S. 502 (1934) (upholding price control regulations in the milk industry as consistent with the rational basis review required by due process).

https://doi.org/10.1017/S1352325201072019 Published online by Cambridge University Press

commercial speech must satisfy an intermediate level of review. Nonetheless the government may choose to regulate commercial speech selectively. For instance, the government can choose to regulate solicitation by lawyers differently from the way it regulates solicitation by accountants.[73] Much closer to hate speech regulation are the content-specific restrictions on speech embodied in employment discrimination law. Section 703 of Title VII of the Civil Rights Act of 1964 prohibits discrimination in conditions of employment on the basis of race, religion, national origin, and gender.[74] Courts have held that harassment by an employer or co-worker that creates a hostile environment counts as discrimination under Title VII for which the employer can be held liable.[75] Like a restriction on hateful fighting words, employment discrimination law targets within the class of workplace harassment only harassment on the basis of race, religion, national origin, and gender.[76] As these examples illustrate, the Court has upheld selective content-specific restrictions within a class of non-fundamental liberties.

Perhaps the government cannot be arbitrary in the subset of unprotected speech it chooses to regulate. Rational basis review, which is the standard of review appropriate to restriction of non-fundamental liberties, demands this. But there is no reason to think that focus, within the class of fighting words, on hateful fighting words is arbitrary. Hate speech is a familiar form of fighting words with a distinctive history, which helps explain why hate speech is a commonly occurring form of fighting words especially likely to cause injury or incite disturbance.

Indeed, content-specific restrictions on hateful fighting words could

73. *Compare* Ohralik v. Ohio State Bar, 436 U.S. 447 (1978) (upholding a disciplinary rule prohibiting lawyers from accepting employment as the result of offering unsolicited legal advice) *with* Edenfield v. Fane, 507 U.S. 761 (1993) (invalidating a state provision prohibiting CPAs from uninvited professional solicitations).

74. 42 U.S.C. § 2000e-2.

75. Meritor Savings Bank v. Vinson, 477 U.S. 57 (1986) (holding that Title VII prohibits sexual harassment that creates a hostile environment); and Harris v. Forklift Systems, Inc., 510 U.S. 17 (1993) (holding that a hostile environment claimant need not show personal injury provided a reasonable person would find the work environment hostile).

76. Scalia tries to distinguish *R.A.V.* and Title VII issues by arguing that selective content-specific regulation of low-value speech is permissible if it produces harmful conduct as "secondary effects." Discriminatory speech in the workplace, he alleges, has the secondary effect of workplace harassment (*R.A.V.* at 389). But in neither the hate speech nor the Title VII context is it especially plausible to see the harm as a remote or secondary effect of the speech in question. In the Title VII context, a hostile workplace environment is not some remote effect of discriminatory speech whose occurrence depends upon myriad other factors; rather, patterns of discriminatory speech themselves constitute a hostile environment. Similarly, as *Chaplinksy* claims, fighting words are words "which by their very utterance inflict injury or tend to incite an immediate breach of the peace" (*Chaplinsky* at 571–72). In any case, whatever we say about the status of the effects of discriminatory speech, there appears to be no disanalogy or asymmetry between *R.A.V.* and Title VII. For the injuries caused by hateful fighting words stand to those words as the hostile environment caused by discriminatory speech in the workplace stands to that speech. So if Scalia agrees that the content-specific regulation of speech contained in employment discrimination law under Title VII is permissible, then he has no basis for denying that the St. Paul ordinance is permissible if, as he concedes, it must be interpreted as restricted to fighting words.

https://doi.org/10.1017/S1352325201072019 Published online by Cambridge University Press

meet Scalia's demand. He claims that even if the entire category of fighting words is unprotected speech, one cannot restrict only a proper part of this category on the basis of its content unless one can show that the very reasons that make the class unprotected in the first place have special application to the proper subclass. But we can provide such a rationale. As we have seen, the reason fighting words count as low-value speech is that they do not engage the deliberative values underlying central First Amendment liberties and are by their nature likely to cause injury or incite disturbance. Whereas all fighting words wound or invite confrontation by being provocative, hateful fighting words are special. Going out of one's way to accost a black student and say "nigger go home" is different from calling someone a "rotten bastard" in the heat of a political argument or demonstration. Hateful fighting words are not simply provocative; invoking historical patterns of systematic discrimination and often violence, they brand their targets as inferior and unworthy of equal concern and respect, and often produce reasonable fears for the target's privacy or safety. In these ways, hateful fighting words are especially traumatic and divisive. As a result, hateful fighting words tend to silence their targets and to undermine the culture of mutual respect and tolerance necessary for a proper representation and fair assessment of alternative points of view in ways that other fighting words don't. If so, hateful fighting words have a distinctive adverse impact on the deliberative values that underlie central First Amendment liberties and that explain why fighting words are low-value speech.[77]

Finally, it's worth noting problems with Scalia's application of strict scrutiny. If these arguments are right, strict scrutiny is the wrong standard of review to apply to content-specific restrictions on hate speech; they should be subject to some less demanding standard of review, such as rational basis review. However, there is no reason to think that restrictions on hate speech could not satisfy strict scrutiny. Scalia concedes that the city of St. Paul has a compelling interest in ensuring the basic human rights of groups that have historically been subject to discrimination.[78] He goes on to ask whether the restriction in question is reasonably necessary to achieve the city's compelling interest and concludes that it is not. "An ordinance not

77. This defense of selective subclassification within the larger category of fighting words suggests a defense of hate speech regulation as an instance of selective subclassification within the larger category of discriminatory speech. Other forms of discriminatory speech may also produce the sort of effects that make hate speech low-value and harmful. But even if the state could adopt more sweeping restrictions on discriminatory speech, that does not establish anything impermissible about the state focusing on the salient subclass of discriminatory speech that employs fighting words. Moreover, as we have seen, there is good reason to think that the very reasons that would make discriminatory speech proscribable apply with special force to the proper subclass of discriminatory speech that employs fighting words. It seems plausible that hate speech's use of traditional epithets or symbols of derision to vilify makes it, in comparison with merely discriminatory speech, more likely to cause emotional distress, to evoke visceral rather than articulate response, and to stifle the expression of marginalized points of view or prevent them from receiving a fair hearing.

78. *R.A.V.* at 395–96.

https://doi.org/10.1017/S1352325201072019 Published online by Cambridge University Press

limited to the favored topics, for example, would have precisely the same beneficial effect."[79] But this is confused. A restriction of all fighting words would serve the interest of protecting members of groups that have historically been subject to discrimination no better than a restriction to hateful fighting words, but the restriction of all fighting words would certainly be more, not less, restrictive. Perhaps Scalia means that a restriction of all fighting words would be the least restrictive means that is not content-specific. Though perhaps true, this claim would be irrelevant. For the question that we are entertaining, if only for the sake of argument, is whether a regulation that is conceded to be content-specific can *nonetheless* pass strict scrutiny. But then it cannot be part of the test that strict scrutiny imposes that the list of means from which the least restrictive one is to be selected include only content-neutral means.

If so, the only good objection to the St. Paul city ordinance is the objection of the concurring opinions that it is over-broad and is not restricted to fighting words. But then *R.A.V.* is no objection to the Stanford provision, because it is so restricted.

XI. HATE SPEECH REGULATION IN PERSPECTIVE

The Millian perspective explains the importance of liberties of expression in terms of deliberative values that are essential to our being morally and politically responsible agents, and some such deliberative rationale seems to guide the Court's own account of central First Amendment liberties. When First Amendment liberties are understood in terms of deliberative values, a plausible case can be made for narrowly crafted restrictions on hate speech. Deliberative values play two roles in the Millian perspective on hate speech. First, deliberative values explain why we should not treat hate speech as high-value speech or recognize freedom to engage in hate speech as a fundamental liberty. Hate speech does not just express discriminatory attitudes; it wounds, threatens, and vilifies its audience in ways that evoke visceral rather than articulate response. Moreover, it undermines the background culture of mutual respect essential for the proper representation and evaluation of diverse points of view, and so discourages participation in a deliberative community. Insofar as hate speech retards rather than advances deliberative values, it is not high-value speech; as a result, the regulation of hate speech need not satisfy strict scrutiny but only some weaker standard of review, such as rational basis review. Second, deliberative values figure in satisfying the relevant standard of review. It is our interest in preventing emotional distress and protecting deliberative interests, especially for those who are already disadvantaged and objects of discrimination, that justifies restricting the freedom to engage in hate speech. This is clearly sufficient to satisfy the relevant standard of review if there is no fundamen-

79. *Id.* at 396.

https://doi.org/10.1017/S1352325201072019 Published online by Cambridge University Press

tal liberty to engage in hate speech. But these considerations about the kinds of harm that hate speech causes and the inequitable way in which they are distributed explain why a good case can be made for thinking that even if hate speech were high-value speech, narrowly crafted hate speech regulations could satisfy strict scrutiny. For it is reasonable to suppose that there is a compelling interest in preventing some citizens from inflicting emotional distress on others, in maintaining a climate of mutual tolerance and respect in which each can deliberate effectively about both private and public matters, and in ensuring a fair distribution of the costs of maintaining a system fundamental liberties. Any hate speech regulation would have to satisfy the least-restrictive-means test, but something very much like the hate speech provisions we have discussed would seem to meet this constraint. Insofar as the Millian perspective on hate speech justifies regulation by appeal to deliberative values, it appeals to the very same values that explain why censorship is normally impermissible. If so, hate speech regulation is a well-motivated exception to the usual prohibition on censorship.[80] Assessing the moral and constitutional dimensions of hate speech and its regulation by Millian principles allows us to put the significance of debates over hate speech into helpful perspective.

For one thing, attempts to regulate hate speech do not have the cultural significance that some like to suppose. It is profoundly confused to see proposals to regulate hate speech as a campaign on behalf of "political correctness" or some other liberal agenda. Hate speech is only a proper part of genuinely discriminatory speech, and a fairly small part, at that. Because politically incorrect speech is presumably an even broader class of speech than genuinely discriminatory speech, it is absurd to suppose, as many have, that hate speech regulations reflect a concern for political correctness. The rationale for hate speech regulation is not to establish, by force if necessary, a liberal consensus in speech and thought. Quite to the contrary, by curbing the most thoughtless, harmful, and divisive expressions of discriminatory attitudes, hate speech regulations aim to establish the kind of culture in which vigorous discussion and debate among disparate views can be both possible and profitable.

Moreover, it would be a mistake to overestimate the practical significance of this defense of hate speech regulation. Hate speech regulation of the sort discussed here is no panacea for racism, sexism, homophobia, or other forms of discrimination and inequality. Though discriminatory speech has discriminatory effects, it is a symptom of underlying discriminatory attitudes, which are themselves symptoms of underlying social relations of inequality and discrimination. If so, hate speech regulation targets symptoms, not the disease. Moreover, many forms of discriminatory speech will

80. If so, Post, in *Free Speech and Religious, Racial, and Sexual Harassment,* is wrong to think that appeal to deliberative values associated with democratic self-governance must condemn narrow restrictions on hate speech. Indeed, the very values on which Post bases his skepticism about the regulation of hate speech provide a plausible rationale for such regulations.

https://doi.org/10.1017/S1352325201072019 Published online by Cambridge University Press

not count as hate speech and so will not fall within the scope of legitimate hate speech regulations. So it would be naïve to suppose that hate speech regulations could play a major role in the battle against discrimination. Nonetheless, hate speech regulations can have significant symbolic value by curbing the most extreme expressions of discriminatory attitudes. Moreover, the existence and recognition of hate speech regulations can serve as a reminder to each member of the community that protected liberties are conditioned on recognizing other members of the community as members in good standing who are entitled to elementary signs of tolerance and respect.

Part of the significance of the Millian perspective on hate speech is jurisprudential. The Millian perspective steers between two methodological extremes sometimes found in the hate speech literature. On the one hand, proponents of hate speech regulation sometimes invoke quite novel interpretive claims that challenge familiar interpretive methods or doctrinal principles. For example, Charles Lawrence embeds his eloquent defense of hate speech regulations within a very controversial reading of Brown v. Board of Education[81] as a First Amendment case.[82] Such views threaten to make the case for hate speech regulation hostage to these novel methods and principles. On the other hand, opponents of hate speech regulation sometimes appear to wield constitutional doctrine and case law in unreflective ways, for instance, when they invoke the increasingly narrow construction the Court has put on the fighting-words doctrine without properly considering whether the reason for treating fighting words as low-value in the first place calls for a broader construction of the fighting-words doctrine.[83] Such critics make the case against hate speech regulation unprincipled. The Millian perspective aims to avoid both tendencies. By invoking deliberative values whose philosophical appeal and importance can be

81. Brown v. Board of Education of Topeka, 347 U.S. 483 (1954) (invalidating policies of segregated education as inherently unequal and, hence, as a violation of equal protection of the laws).

82. Lawrence interprets the state policy of school segregation as expressive activity and claims that *Brown* should be read as a First Amendment case in which the Court recognizes the permissibility of regulating discriminatory speech. If so, then *Brown* becomes a precedent for upholding hate speech regulation. But it is implausible, as well as unnecessary, to read *Brown* as a First Amendment case. Lawrence anticipates resistance to his interpretive claim based on the state-action doctrine—which claims that individual rights recognized in the Bill of Rights and the Fourteenth Amendment are claims against state action, not claims against the actions of private persons—and challenges that doctrine. But the state-action doctrine is a red herring. The problem with viewing *Brown* as a First Amendment case is not that this gives individuals constitutional rights against other individuals as well as the state; rather, the problem is that it gives rights to the state as well as to individuals. In other words, whereas the state-action doctrine concerns *whom one has rights against*, the obstacle to reading *Brown* as a First Amendment case concerns *who has the rights*. It seems quite clear that the Bill of Rights and the Fourteenth Amendment give rights to individuals, not to the states (and, according to the state action doctrine, these are rights against state action). So I think that Lawrence's novel interpretive claim about *Brown* has very little to recommend it. Fortunately, neither *Brown* nor the defense of hate speech regulation requires us to take on this interpretive baggage.

83. *See* Strossen, *Regulating Racist Speech on College Campus*, at 515.

https://doi.org/10.1017/S1352325201072019 Published online by Cambridge University Press

explained in a straightforward way and that underlie central aspects of First Amendment jurisprudence, I have tried to explain the legitimacy of narrowly crafted hate speech regulations. Insofar as hate speech retards rather than advances deliberative values, hate speech regulation can be defended by appeal to the very same values that explain why content-specific censorship is normally impermissible. If so, hate speech regulation represents a principled exception to the usual prohibition on content-specific censorship.

Finally, the Millian perspective on hate speech draws our attention to important but difficult questions in political and constitutional theory about how one determines which interests and liberties are fundamental ones to which individuals can claim rights. It is generally recognized that individuals do not have rights to liberty per se, if only because the civil and criminal law legitimately restricts some liberties in order to prevent various kinds of nuisance and harm. We have rights not to liberty as such, but to important or fundamental liberties, including First Amendment liberties. Constitutional theory recognizes this; it treats some interests and liberties as more fundamental than others. It recognizes these fundamental interests and liberties as constitutional rights by according them special protection; it subjects legislation that interferes with fundamental interests and liberties to strict scrutiny or some comparable standard. Though constitutional theory does treat First Amendment liberties as fundamental liberties, it does not recognize a First Amendment liberty to speech as such; First Amendment jurisprudence distinguishes between low-value and high-value speech and recognizes central First Amendment protection only for high-value speech. To explain which liberties, including which liberties of expression, are fundamental and why, one must apparently invoke values other than liberty itself. The Millian perspective explains the importance of liberties and other interests in terms of deliberative values that are essential to our being morally and politically responsible agents, and deliberative values explain the Court's own account of central First Amendment liberties. On this view, fundamental liberties are explanatorily posterior to deliberative values. When fundamental liberties are understood in deliberative terms, I have argued, a plausible case can be made for refusing to recognize a fundamental liberty to engage in hate speech. Insofar as hate speech retards rather than advances deliberative values and undermines the background culture of mutual respect essential for the proper representation and evaluation of diverse points of view, hate speech regulations should not be seen as restricting fundamental liberties. Furthermore, if the importance of fundamental liberties implies that we must be prepared to pay a price for recognizing rights to those liberties, then the costs ought to be distributed equally or fairly. Insofar as the costs of including hate speech among central First Amendment liberties are borne disproportionately by already marginalized groups, we should reject that interpretation of fundamental First Amendment liberties.

https://doi.org/10.1017/S1352325201072019 Published online by Cambridge University Press

This gives the Millian perspective on hate speech much wider jurisprudential significance. For we can ask what deliberative values imply about other First Amendment issues—including the permissibility of regulating pornography and obscenity, the appropriate level of scrutiny for legislation affecting commercial speech, the legitimacy of a fairness in broadcasting doctrine that requires media to devote broadcast time to issues of public importance and to represent diverse viewpoints on broadcast topics, and the permissibility of restrictions on political speech involved in campaign finance reform—and about the correct interpretation of other fundamental interests and liberties—including the nature and scope of a right to privacy. These are all issues on which the Millian perspective, articulated here, is well worth exploring.[84] These explorations must be reserved for another occasion. The present discussion suggests that the Millian perspective would be a valuable guide.

84. There will be overlap between the Millian perspective on these issues and the perspective rooted in democratic self-governance. *Cf.* SUNSTEIN, DEMOCRACY AND THE PROBLEM OF FREE SPEECH, and OWEN FISS, THE IRONY OF FREE SPEECH (Cambridge, MA: Harvard University Press, 1996).

https://doi.org/10.1017/S1352325201072019 Published online by Cambridge University Press