

ARTICLE

John Stuart Mill's Harm Principle and Free Speech: Expanding the Notion of Harm

Melina Constantine Bell 

Washington and Lee University
Email: bellm@wlu.edu

Abstract

This article advocates employing John Stuart Mill's harm principle to set the boundary for unregulated *free* speech, and his Greatest Happiness Principle to regulate speech outside that boundary because it threatens unconsented-to harm. Supplementing the harm principle with an offense principle is unnecessary and undesirable if our conception of harm integrates recent empirical evidence unavailable to Mill. For example, current research uncovers the tangible harms individuals suffer directly from bigoted speech, as well as the indirect harms generated by the systemic oppression and epistemic injustice that bigoted speech constructs and reinforces. Using Mill's ethical framework with an updated notion of harm, we can conclude that social coercion is not justified to restrict any harmless speech, *no matter how offensive*. Yet certain forms of speech, such as *bigoted insults*, are both harmful *and* fail to express a genuine opinion, and so do not deserve free speech protection.

In 1994 Stanford University sought to prohibit on campus, as discriminatory harassment: “personal vilification of [individuals] on the basis of their sex, race, color, handicap, religion, sexual orientation, or national or ethnic origin.” Under this policy, speech would be regarded as personal vilification only if it met each of these criteria: (1) “is intended to insult or stigmatize an individual or small group of individuals on the basis of” one of the listed categories; (2) “is addressed directly to the individual or individuals whom it insults or stigmatizes”; and (3) “makes use of insulting or ‘fighting words’ or nonverbal symbols,” which “tend to incite to an immediate breach of the peace,” and which are commonly understood to convey direct and visceral hatred or contempt for human beings on the basis of’ the listed categories. A California court struck down Stanford’s policy on grounds that it violated students’ free speech rights.

The American Civil Liberties Union articulates the usual concern with speech policies:

How much we value the right of free speech is put to its severest test when the speaker is someone we disagree with most. Speech that deeply offends our morality

or is hostile to our way of life warrants the same constitutional protection as other speech because the right of free speech is indivisible: When we grant the government the power to suppress controversial ideas, we are all subject to censorship by the state. Since its founding in 1920, the ACLU has fought for the free expression of all ideas, popular or unpopular. Where racist, misogynist, homophobic, and transphobic speech is concerned, the ACLU believes that more speech – not less – is the answer most consistent with our constitutional values.¹

That is the civil liberties side of it. But the civil rights side, advocated by Richard Delgado, Mari Matsuda, Charles Lawrence III, Kimberlè Crenshaw and others, takes a different view. For them, absence of government speech regulation grants private parties with the most social capital expansive freedom to express themselves. The powerful are free to speak, and also free to silence the vulnerable by intimidating them, and by reinforcing inaccurate stereotypes that undermine their credibility. Thus, even if those belonging to historically stigmatized groups attempt to speak (which may require them to risk an important social alliance, a job, or even their personal safety), they may end up with no audience or with an audience that sharply discounts their testimony. Moreover, the dominant may engage in expression that causes psychological harm through overt intimidation and/or targeted insults that function as verbal assaults. Harm to physical and mental health occurs when stress levels are perpetually elevated by living in a constant state of hyper-vigilance.² Concrete, measurable financial disadvantages can accrue as a result of intentional and inadvertent discrimination. Forms of expression, including media portrayals, bigoted humor, and verbal reinforcement of stereotypes in daily conversation, contribute to biased attitudes and a hostile environment for historically stigmatized groups.

How might a society or institution safeguard the free exchange of ideas, for the sake of discovering truth and fostering individual freedom, while simultaneously ensuring the equitable distribution of testimonial credibility, opportunities for free expression, and protection from unjustified harm? In this article, I propose a principled framework for striking the right balance among these important competing claims. The centerpiece of this framework is nineteenth-century English philosopher John Stuart Mill's *harm principle*, which rules out regulation of speech based on its offensiveness to particular persons, regardless of the number offended. The harm principle checks majorities who, regarding a view as distasteful or false, would otherwise have the authority to suppress it. It also provides society with an effective means of preventing powerful private parties from harming vulnerable ones, even when their chosen weapon is verbal expression. Mill himself might have favored the civil liberties paradigm because he did not have the benefit of modern scientific knowledge about forms of tangible psychological harm, mechanisms for transmitting cultural norms, implicit bias, structural discrimination, etc. Coupled with a more expansive, modern conception of harm, the principle can provide a valuable resource for separating the speech that should be addressed with more speech from the speech acts that unjustifiably harm others and should be discouraged or sometimes even coercively restrained. Even as prominent and resolute a free speech advocate as Mill could recognize this, within his ethical framework, given currently available facts.

¹ACLU, Speech on Campus <<https://www.aclu.org/other/speech-campus>> [accessed 1 June 2020].

²LaBarron K. Hill and others, Ethnic Differences in Resting Heart Rate Variability: A Systematic Review and Meta-Analysis, *Psychosomatic Medicine* 77.1 (2015), 16–25 (pp. 8–9).

My basic argument is that we should use Mill's harm principle to set the boundary for unregulated *free* speech, and Mill's Greatest Happiness Principle (GHP) to regulate speech that threatens unconsented-to harm. Speech should be regulated to minimize such harm, ensuring that the regulation itself does not cause greater harm than it prevents, and that the interests of all persons are treated with equal concern. Using these principles, I hope to show that (1) harmless speech – no matter how offensive it may be – should not be suppressed coercively; but (2) certain forms of speech – such as *bigoted insults*, which fall into a category of what is sometimes called 'assaultive speech' – are harmful and fail to express a genuine *opinion* or *viewpoint*, and so do not deserve free speech protection.

In Part I, I provide an account of how Mill envisioned his harm principle safeguarding free discussion, using Joel Feinberg's influential account of the distinction between harm and offense. In Part II, I consider arguments and evidence that demonstrate the need to broaden the concept of harm beyond Mill's understanding of it. In Part III, I use the harm principle thus expanded and the GHP to explain why the rationale for free speech does not justify protection of bigoted insults, even though it justifies protecting the content of offensive opinions and viewpoints that may accurately be categorized as bigoted. I conclude that it is not morally justifiable to use coercive means (including formal legal sanctions and the informal social sanction of *moral disapprobation*) to attempt to silence or restrain merely offensive speech, although some harmful speech that has gone unregulated can, with moral justification, be subject to social regulation. I focus on bigoted insults as an example of harmful speech that has been protected without justification. Using the Millian framework I propose, society is likely justified in restricting other forms of harmful speech that it now accepts, defends, and protects.

I. Mill's harm principle and free speech

Recognizing the possibility of *tyranny of the majority* within a democracy, especially considering how strongly people tend to believe that their own views are correct, Mill wrote his influential essay *On Liberty* to articulate and defend the *harm principle*:

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.³

As Mill explains, the harm principle is a jurisdictional principle that sets a boundary for social coercion, whether it takes the form of law or informal social sanctions.⁴ Mill is a proponent of liberalism, a political philosophy that entails a presumption in favor of individual liberty, and requires that any limitation of liberty be justified by a liberty-limiting principle. A liberty-limiting principle identifies a *prima facie* reason to justify social coercion, and justifies that coercion in the absence of stronger countervailing reasons.⁵ For Mill, the GHP and the harm principle are the only

³Mill, *On Liberty*, CW XVIII, pp. 223–24. Citations of Mill marked by 'CW volume number, page number' refer to the *Collected Works of John Stuart Mill*, ed. John M. Robson, 323 vols (Toronto: University of Toronto Press, London: Routledge and Kegan Paul, 1963–91).

⁴*Ibid.*, pp. 224–25.

⁵Joel Feinberg, *The Moral Limits of the Criminal Law*, 4 vols (Oxford University Press, 1984), I, pp. 9–10. Mill, unlike Feinberg, would have to regard the GHP as a liberty-limiting principle because promoting

acceptable liberty-limiting principles. Unless a person's conduct causes a definite risk of harm to other persons, who either do not consent to the risk or who lack the capacity to consent, that conduct is outside society's jurisdiction. All coercive social paternalism aimed at competent adults is foreclosed. Additionally, coercive social regulation of conduct that merely offends (without harming) others is unjustified.

By *social coercion*, I mean to include both legal penalties and moral disapprobation. In both cases, social force or power is aimed at an individual to cause them to change their behavior in order to avoid the pain of social disapproval and/or exclusion. This is to be contrasted with educating someone,⁶ remonstrating with them,⁷ and helpfully alerting them to possible consequences.⁸ The latter activities appeal to a person's reason and autonomy: they aim to convince a person of a view or equip them with useful information they can use to make their own decision. Social coercion, as I use it, motivates someone to act differently than they would in its absence to avoid pain, rather than because they recognize reasons to act differently and choose to alter their course.

Because the harm principle places within society's jurisdiction to regulate only those actions that pose a significant risk of harm to individuals other than the competent actor, a clear distinction between conduct that is harmful and conduct that is disagreeable, though harmless, is critical. In *The Moral Limits of the Criminal Law*, Joel Feinberg attempts to draw this distinction, first setting aside the sense of 'harm' that means any sort of damage to any object; this is not the relevant sense for Feinberg, Mill, or me. The harm that concerns us is that which constitutes "a setback to the interests" of a person. *Interests* are a person's stake in certain matters, such that their life goes better or worse depending on how these matters develop. Feinberg illustrates this with some examples: a person has an interest in the advancement of their career, the preservation of their property, the economic and social conditions of their country, and the well-being of family and people they love. But the most ardent fan of a baseball team is not harmed by loss of even the most important game. Disappointments of this sort do not set back a person's interests. Experiences of temporary discomfort or pain – "unhappy mental states" – do not (as such) set back interests. If they pass and leave a person as they were, whole and undamaged, their interests were not set back.⁹ Feinberg further clarifies the distinction: "[A]n affront or an insult normally causes a momentary sting; we wince, suffer a pang or two, then get on with our work, unharmed and whole. But if the experience is severe, prolonged, or constantly repeated, the mental suffering it causes may become obsessive and incapacitating, and therefore harmful."¹⁰

In my view, when Mill refers to 'harm' he has in mind a broader concept than Feinberg, who refers exclusively to setbacks caused by one person violating another's rights.¹¹ For Mill, legal sanctions are only justified to safeguard rights. But for violations of imperfect duties, which lack correlative rights, moral disapprobation may be justified. As a form of social coercion, however, it is justified only to counter a substantial risk of

aggregate happiness is a prima facie reason to justify social coercion (even when the harm principle provides a stronger reason not to do so).

⁶Mill, *On Liberty*, CW XVIII, p. 282.

⁷Ibid., p. 224.

⁸Ibid., p. 277.

⁹Feinberg, I, pp. 31–44.

¹⁰Ibid., pp. 45–46.

¹¹That Mill and Feinberg have different notions of harm should not be surprising, given their different purposes: for Mill, to protect a personal zone of social non-interference; for Feinberg, to determine when *criminal* punishment is justifiable.

harm to nonconsenting others.¹² Because sometimes social coercion is justified though no rights are violated, there must be harm on Mill's view even when there is no harm on Feinberg's.

While I find Feinberg's account of harm too narrow, I find Piers Turner's, which regards offense as a category of minor harm – where harm is any direct negative consequence for others – too broad.¹³ I believe it is critically important to distinguish harm – a tangible, objectively verifiable setback to a person's interests – from offense, which is a subjective reaction to an experience that can vary greatly among persons. I agree with Turner that intense feelings of offense can exceed the painfulness of some harms. For example, losing a dollar from your wallet without ever noticing that it is missing is a very minor setback to interests (harm), while public criticism of your parenting skills is undoubtedly more painful (offense). Turner appears to equate offense with emotional distress, which I would resist. For me, offense is what Feinberg calls "unhappy mental states": temporarily unpleasant, but not causing sustained damage to well-being interests. To know a person is offended, one must know about the person's subjective state: a person cannot be mistaken about whether they are offended. A person *can* be mistaken about whether they are harmed, or unaware of the harm (for example, when a disease caused by environmental pollution or drug side effects has yet to manifest symptoms). But harm is objectively verifiable with access to the relevant situational information. Pain or distress, though a mental state, can cause or constitute objective harm if it is severe, prolonged, and/or repeated. Chronic headaches and post-traumatic stress, for instance, can interfere with normal functioning. So the offense-harm distinction, in my view, is not determined by the intensity of discomfort a person suffers, but by whether it is (or becomes) grounded in an objective state of the person, rather than *solely* residing in a temporary subjective state of the person. To take seriously Mill's concern about protecting minorities from tyranny of the majority, we should categorically reject the possibility of justifying coercion solely to accommodate the sensibilities of those who happen to constitute a majority at a particular time. For that reason, I distinguish offense from harm, and reject using an offense principle as a liberty-limiting principle.¹⁴

Significant wrongful threats to personal safety, bodily and mental health, means of earning a living or securing a place to live, or to the legal rights that safeguard a person's key liberties and opportunities are clearly types of harm. Society is authorized to use coercive force to protect nonconsenting people from these threats to the extent the coercion will not cause a greater harm or threat of harm to individuals in society than the harm whose prevention is sought. This Millian liberal position is to be contrasted sharply with the principle of legal moralism,¹⁵ which regards limitations on individual liberty as justified to protect the majority's sensibilities or moral judgments. In *On Liberty*, this contrasting position is represented by a temperance movement that claims the government is obligated to prohibit liquor sales to protect their "social right" to be free of drunken misbehavior. Mill denounces this theory of social rights, which

¹²Mill, *On Liberty*, CW XVIII, p. 276.

¹³Piers Norris Turner, 'Harm' and Mill's Harm Principle, *Ethics* 124.2 (2014), 299–326. Turner argues that Mill simply intended for his harm principle to be a bulwark against social paternalism.

¹⁴Feinberg, II, p. x. Feinberg defines 'offense' as an objective state of the person because his notion of it limits its causes to moral wrongs and its mental states to "a universally disliked kind" (II, p. 2). I find this account problematic in ways that go beyond the scope of this article.

¹⁵Feinberg also rejects legal moralism as illiberal (IV, p. x).

authorizes coercive regulation to force people to behave as the majority believes they ought, as a “monstrous principle” that effectively nullifies individual liberty.¹⁶ Such a principle gives people too little control over their own behavior and lives, and too much control over others’. Agreeing with Mill on this point, I endorse Mill’s version of liberalism, with the harm principle as the only liberty-limiting principle besides the moral criterion itself (for Mill, the GHP). I believe the harm principle is judged insufficient, when it is, because the scope of harm – understood as setbacks to objectively identifiable interests – is drawn in a biased way to exclude important instances.

For Mill, the harm principle is derived from his GHP¹⁷ and serves as a corollary to it.¹⁸ It is subordinate to the GHP in the sense that its function is to provide a safer route to maximal aggregate happiness by bypassing, in cases of self-regarding or harmless conduct, separate case-by-case calculations to determine whether social restraint of individual liberty is justified. As with the Nautical Almanac,¹⁹ we are relieved from performing numerous calculations and multiplying opportunities to make mistakes. Mill apparently judges that a blanket rule prohibiting social interference in cases of self-regarding and harmless conduct produces greater aggregate happiness than the rule that we perform the utility calculation in each individual case. But even though an action that harms others *is* within society’s jurisdiction to regulate coercively, such regulation should occur only to the degree that it maximizes aggregate happiness. So in making decisions about whether to adopt speech-restrictive policies, first we should determine that the behavior under consideration *does* pose a risk of harm to others and therefore falls within society’s jurisdiction. Only if the behavior falls within society’s jurisdiction may utilitarian calculations take place to determine whether the benefits of regulating the conduct outweigh the harm of doing so. We should implement the policy if and only if the behavior is within society’s jurisdiction *and* the policy’s aggregate benefits outweigh its aggregate harms. Consequently, when blameless conduct poses a risk of harm, it seems unlikely that aggregate utility could be served better by social coercion than by less intrusive means. Social coercion is intrinsically painful to the person(s) coerced, and will only be justified where *necessary* to prevent greater harm to others. Harm that is not intended cannot be deterred by punishment or force more effectively than through education or other social supports, and deterrence would be counterproductive with respect to harm that is an unavoidable byproduct of a socially useful activity benefitting all.²⁰

In light of Mill’s harm principle, consider Mill’s emphatic defense of free expression and discussion in *On Liberty*. Mill acknowledges two important roles free expression plays in human happiness: an individual or private role, and a social or public role. Free expression (including participation in discussion) has what I will call *private value* because it enables individuals to gain knowledge, develop their capacities, and express their individual personalities.²¹ Free expression has what I will call *public value* because, especially in discussion of opinions, it contributes to the accumulation

¹⁶Mill, *On Liberty*, CW XVIII, p. 288.

¹⁷ “[A]ctions are right in proportion as they tend to promote happiness; wrong as they tend to produce the reverse of happiness” for all persons considered collectively. Mill, *Utilitarianism*, CW X, p. 210.

¹⁸ “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.” Mill, *On Liberty*, CW XVIII, p. 224.

¹⁹Mill, *Utilitarianism*, CW X, p. 225.

²⁰Mill, *On Liberty*, CW XVIII, pp. 292–93.

²¹Christopher Macleod seems correct to assign Mill’s defense of freedom of discussion to *On Liberty* Chapter II, and his defense of expressing one’s personality to Chapter III. Macleod, *Mill on the Liberty*

of collective knowledge and to social progress. Free discussion is of both private and public value as a check on government power.²² Individuals privately experience restrictions on these activities as intrinsically painful, so there must be powerful countervailing reasons for them. For Mill, freedom is an “essential” of human well-being, so freedom of opinion *and its expression* are necessary components of individual well-being.²³ However, security is also essential to human well-being.²⁴ So when expressive actions risk harm to others and are within society’s jurisdiction to regulate, there seems to be no *general answer* as to whether private utility will be maximized by permitting these expressive actions or restraining them to some degree. Decisions about whether to regulate particular forms of potentially harmful expression must be made case by case, after carefully weighing the utilities on each side. A pivotal consideration in many cases will be the public value of particular forms of speech. Mill cautions that silencing an opinion is not “simply a private injury,” but is identical with “robbing the human race” of resources vital to the pursuit of truth through discussion and exposure to new ideas.²⁵

Looking ahead to Part III, I will there consider how to balance conflicting private interests and the public interest using a hypothetical example. First, in Part II, I explain how bigoted expression and discrimination can set back a person’s interests in physical and mental health in concrete, measurable ways; and how members of marginalized groups with no opportunity to withhold consent to certain forms of bigoted speech might find their own capacity to speak and be heard greatly curtailed. Their work, housing, personal safety and/or liberty might also be compromised.

II. Expanding the notion of harm

In this part, I want to highlight the *general* point that because bigoted speech causes direct harm to individuals, and plays a central role in maintaining and perpetuating institutional oppression and injustice, it causes not merely subjective unpleasant mental states or offense, but also serious harms that usually remain unrecognized both in law and in public discussion. Even those who are not offended, or those who do not belong to the groups targeted by bigoted speech, can be harmed by it. Although I am not the first to make this argument, I wish to reiterate it in light of recent empirical evidence. My second, more *specific* objective, undertaken in Part III, will be to explore circumstances in which bigoted speech might justifiably be prohibited because it is both harmful, and *also* fails to communicate a viewpoint that serves as a basis for consideration or discussion.

A. Tangible, concrete harms

Members of historically marginalized social groups suffer tangible, concrete harms both directly from exposure to bigoted speech, and indirectly because of the hostile social environment, discrimination, and oppression that result. I will briefly describe

of Thought and Discussion, *The Oxford Handbook of the Freedom of Speech*, ed. A. Stone and F. Schauer (Oxford University Press, forthcoming).

²²Mill, *On Liberty*, CW XVIII, pp. 222–23.

²³Ibid., pp. 257–58.

²⁴Mill, *Utilitarianism*, CW X, pp. 250–51.

²⁵Mill, *On Liberty*, CW XVIII, pp. 228–29.

experiments identifying some of these harms to help convey their concrete and material nature and to demonstrate that they are not, in Feinberg's words, mere "unhappy mental states." Instead, they involve "experience [that] is severe, prolonged, or constantly repeated" such that "the mental suffering it causes may become [. . .] incapacitating, and therefore harmful."²⁶

First, consider sexist humor, which can tangibly and directly harm women by eliciting depression, eating disorders, disruption of focused attention, appearance anxiety, and body shame. It can indirectly harm women by instigating discrimination. In one study, a group of female participants exposed to sexist comedy skits expressed a greater state of self-objectification compared to women exposed to neutral comedy skits, while male participants' ratings of self-objectification did not differ based on which skits they viewed.²⁷ Other experiments found that women engaged in more body surveillance after watching sexist comedy clips compared to women who watched neutral comedy clips.²⁸ The skits were considered sexist if they reduced women to sex objects, depicted them as a caricature of traditional gender roles, or played on specific sexist stereotypes of inferiority. Examples included a skit about "good wife school," a man telling a woman "you're an attractive waste of time" at this party "unless you want to sleep with me," and a Daniel Tosh skit about women being able to do what men can do "except" a long catalog of tasks.²⁹

Women were not *merely* offended by the humor (finding it less funny than men according to ratings). More importantly, they had greater body surveillance, a kind of negative self-focus that can disrupt focused attention, usurping attentional resources and reducing performance on other cognitive tasks. In other research, women performed worse on math tests while wearing swimsuits, relative to a comparison group wearing sweaters. Their attention seems to have been divided between their appearance and the test.³⁰ Lower performance can mean not getting the degree, the job, the raise, or the promotion. It can have concrete effects on mental health and material circumstances.

Sexist humor has greater capacity to harm women than straightforward sexist statements. Humor places women's "disadvantaged status in a more acceptable light and thus bypasses criticism and rejection." Sexist jokes imply that it is acceptable to make demeaning statements about women aloud, reinforcing notions that it is culturally acceptable to regard women as unequal. "Humor cues people to suspend their usual critical or literal ways of responding to a message" and "creates the perception that its underlying message is socially acceptable." Therefore, women who hear these jokes might internalize their messages and self-objectify rather than rejecting or criticizing the jokes. Self-objectification, in turn, has been experimentally linked to depression, eating disorders, disruption of focused attention, appearance anxiety, and body shame.³¹

Besides harming women directly, sexist humor can activate discrimination against women. One study asked men who expressed sexist attitudes ("sexist men") to respond

²⁶Feinberg, I, pp. 45–46.

²⁷Thomas E. Ford and others, Sexist Humor as a Trigger of State Self-objectification in Women, *Humor* 28 (2015), 253–69, at p. 259.

²⁸Ibid., pp. 261–64.

²⁹Ibid., pp. 257–59.

³⁰Ibid., p. 265.

³¹Ibid., pp. 255–65.

to a vignette about a workplace interaction, and found that prior experimental “exposure to sexist jokes led to greater tolerance of the supervisor’s sexist behavior in comparison to exposure to neutral jokes or comparable non-humorous sexist statements.”³² Additionally, after watching sexist comedy skits, sexist men demonstrated significantly greater willingness to cut funding for a women’s organization, compared to other types of organizations. This was not true of sexist men who watched neutral comedy skits.³³

Mill recognizes that writing a newspaper opinion that corn merchants are starvers of the poor is an importantly different matter from shouting the same opinion to an angry mob gathered at the merchant’s business or home: “even 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.”³⁴ As described above, experimental evidence suggests that expressing views such as “women only have value as wives or sexual partners” is significantly less likely to cause harm to women than expressions of comparable views using humor, *which bypasses rational consideration, as does an incendiary statement to an angry mob*. Jokes do not contribute to discussion as propositions with truth values do,³⁵ and therefore jokes have lower social value. So while opinions should be protected, the manner of expressing those opinions could be shaped to optimize their potential contributions to rational discussion while minimizing their likelihood to harm socially vulnerable, disadvantaged, or oppressed people, whether the foreseeable harm is imminent or festering, caused simply or by aggregation of contributory causes.

As we have seen, sexist humor can cause serious harm. Now consider how African Americans might be tangibly harmed by racist speech. A 2018 overview of recent psychology studies explores the relationship between discrimination and physical health in African Americans.³⁶ The authors state explicitly that they focused on the link between discrimination and objective physical health outcomes because those are “more mechanistically linked to disease and mortality than mental health and other self-reported outcomes.” On major indicators of health (coronary heart disease, cancer, stroke, HIV) African Americans do worse in the US than other ethnic groups, even controlling for socioeconomic status (SES), health behaviors, and access to care. Researchers believe this may be due to *discriminatory stressors*. African-American men and women often experience different sorts of discrimination. For example, most often men experience racial profiling and mistreatment by police, and women experience interpersonal incivilities and wage discrimination. Black women have a stronger association between discrimination and health outcomes compared to Black men, perhaps due to intersectional discrimination (involving interactions between race and gender oppression). Regarding SES, at “least one study has found that the effects of discrimination on cortisol (a physiological marker of stress activation) were more harmful for African Americans with higher versus lower levels of education.” Low SES has, independently of race/ethnicity, been linked to worse health outcomes; for example, it is associated

³²Julie A. Woodzicka and Thomas E. Ford, A Framework for Thinking about the (not-so-funny) Effects of Sexist Humor, *Europe’s Journal of Psychology* 6.3 (2010), 174–95 <<https://doi.org/10.5964/ejop.v6i3.217>> (p. 182).

³³Ibid., p. 185.

³⁴Mill, *On Liberty*, CW XVIII, p. 260.

³⁵Christopher Macleod, Truth, Discussion, and Free Speech in *On Liberty* II, in press.

³⁶Tené T. Lewis and Miriam E. Van Dyke, Discrimination and the Health of African Americans: The Potential Importance of Intersectionalities, *Current Directions in Psychological Science* 27.3 (2018), 176–82 (p. 176).

with a “C-reactive protein (CRP), a marker of inflammation linked to later disease.” Yet SES discrimination is associated “with higher levels of CRP among college-educated African Americans, but not among African Americans without a college degree or Whites at any level of education.” This appears to indicate that harmful stress activation in African Americans is not reducible to SES discrimination; racial discrimination itself can cause direct physical harm in the form of tangible adverse health outcomes.³⁷

The study’s authors define “discrimination” as “the unjust or prejudicial treatment of a person or group of people, particularly on the grounds of characteristics such as race, age, or gender,” and their examples involve discrimination in housing, policing, and medical care.³⁸ However, it would be surprising if racist jokes and insults, which fall within this broad definition of discrimination, failed to affect African Americans in similar ways, especially given the demonstrated connection between sexist humor and harm to women. Instances of racist or racially insensitive speech (such as microaggressions) are likely to be more pervasive and occur more regularly in a person’s life than the other forms of discrimination mentioned. And self-reported experiences of discrimination have already been linked to worse health outcomes across various groups marginalized on the basis of their race, gender, SES, age, and/or sexual orientation.³⁹

B. Systemic oppression and epistemic injustice

In the US, where free speech is stringently protected, the civil liberties view is most familiar: the proper response to speech that is erroneous, incomplete, bigoted, etc. is more speech, and government should not suppress speech or declare truth. The civil rights view, however, draws attention to oppressive social background conditions, which lead to starkly unequal social capital to be leveraged in the celebrated *marketplace of ideas*. The civil rights view needs more emphasis and attention in considering what constitutes genuine harm and its magnitude.

Constitutional interpretation unfolds in the US case by case, and with it public conceptions of free speech. A state actor’s attempt to restrict speech can be understood in isolation from other happenings and conditions within society: a KKK leader is convicted of involvement in a criminal syndicate for making a speech at a rally;⁴⁰ public high school students are expelled for wearing black armbands in protest against the Vietnam War.⁴¹ US courts typically speak as though their adjudications simply apply relatively straightforward rules to sets of facts in a syllogistic manner. This obscures, perhaps even from the judges themselves, the way decisions serve the interests of dominant social groups. Most judges and members of society who do not belong to a marginalized demographic category interpret bigoted utterances as isolated events too. They are not seen for what they are: part of an oppressive system of unfair discrimination in employment, exclusion from housing and educational opportunities, police violence and mass incarceration, etc. Members of oppressed groups are not the people who made legal rules to fit their experience and protect their interests, so often their experiences and interests are left out of the systems designed to protect people from, and compensate them for, others’ encroachments. White privilege and male privilege, for

³⁷Ibid., pp. 179–80.

³⁸Ibid., p. 176.

³⁹Ibid., p. 180.

⁴⁰Brandenburg v. Ohio, 395 U.S. 444 (1969).

⁴¹Tinker v. Des Moines Independent Community School District, 393 U.S. 503 (1969).

example, make harm that is only suffered by those who are marginalized difficult to see, and the legal framework in place for hundreds of years further obscures it.

Bigoted utterances are among, to borrow Marilyn Frye's metaphor,⁴² the wires in a birdcage that cut off movement in every direction, pressing the groups it contains between inescapable barriers. Looking at a single wire, one might wonder why the bird does not simply fly around it to escape. One must see the system as a whole before one can grasp that it is oppression. When considering particular instances of bigotry, such as a racist epithet shouted at a person, it is common for members of dominant groups in society to see an isolated event that can be shrugged off. It is easy to miss how such insults play a role in reproducing and reinforcing an oppressive system in which members of particular social groups are vulnerable because of their group identities. The harm caused by bigoted insults is possible only because of the system in which they play a role. There are insults against people of color and LGBTQ+ people, for example, that trade on a particular history of discrimination, so there is no equivalent way to insult, intimidate, stigmatize or exclude a person for being white, heterosexual, or cisgender. The weight of historical contempt and implicit bias cannot be leveraged against members of dominant groups in the way it can against members of marginalized groups. The invisibility of hierarchical social oppression to most members of dominant groups conceals from them how members of marginalized groups are asymmetrically vulnerable to group-identity insults.

Miranda Fricker develops a useful conceptual framework for understanding why the costs and benefits of free speech are not equitably distributed: the problem of epistemic injustice.⁴³ Epistemic injustice occurs when a discriminatory social attitude toward a person limits that person's ability to participate in social practices of sharing knowledge and exchanging ideas. Fricker identifies two different kinds of epistemic injustice: testimonial injustice and hermeneutical injustice. *Testimonial injustice* limits the person's participation because, due to their identity, their word is not regarded as credible. *Hermeneutical injustice* limits a person's participation because their experiences are difficult to articulate in a way that is intelligible, perhaps even to them, given the social concepts and language available. Their perspective is thus marginalized, obscured from the common store of knowledge.

Fricker illustrates testimonial injustice by discussing the situation of Tom Robinson, the African-American defendant in Harper Lee's *To Kill a Mockingbird*, who is accused of raping a white girl, Mayella. In fact, Mayella tried to kiss him when he stopped to help her with some household chores, and he ran. If he had pushed her away, he would no doubt be accused of assault; if he did not, he would also be accused of assault. That is the double bind of oppression. Running away also made him appear guilty. So he could not tell the truth on the witness stand and sound credible. When asked why he would help her with chores for free, he answered honestly that he felt sorry for her, because she had no one to help her. But the statement that a Black person felt sorry for a white person was incomprehensible and perceived as impertinent by the courtroom audience, who could not believe Robinson, who was Black, could be motivated by human sympathy. Besides, there was already a stereotype working against Robinson that Black people tend to lie. The jury's epistemic failure – failure to perceive

⁴²Marilyn Frye, *The Politics of Reality: Essays in Feminist Theory* (New York: The Crossing Press, 1983), pp. 2–5.

⁴³Miranda Fricker, *Epistemic Injustice: Power and the Ethics of Knowing* (Oxford: Oxford University Press, 2007).

Robinson as fully human or as a truthful witness – led to a terrible ethical failure, as they found an innocent man guilty.⁴⁴ Fricker labels this type of epistemic injustice an “identity-prejudicial credibility deficit” because the speaker’s identity makes the hearer judge his testimony as significantly less truthful than it is. And social attitudes and dispositions toward the stereotyped group tend to be persistent and systematic, rather than incidental.⁴⁵

Hermeneutical injustice arises when the prevailing social framework lacks the concepts or language to describe a particular experience, placing members of marginalized groups at a disadvantage in trying to articulate their experiences and advocate improvements. Fricker offers two historical examples: postpartum depression and sexual harassment. Women experienced both for a long time without a vocabulary or conceptual resources to understand or explain what they were enduring. This can make members of marginalized groups feel disconnected from reality and alone.⁴⁶

III. Striking the right balance

In Part II, I demonstrated that bigoted speech can result in harm, in Feinberg’s sense of a setback to the interests of persons, contrasted with merely subjective unpleasant mental states or offense. This brings it within society’s jurisdiction to regulate under Mill’s harm principle. In this part, I argue that *bigoted insults*, a category of bigoted speech, may justifiably be prohibited because in addition to being harmful, they *also* fail to communicate a viewpoint that serves as a basis for consideration or discussion. From the standpoint of the Greatest Happiness Principle, because they do not benefit humans as progressive beings by contributing to discussion, they have no public value and produce no benefit that can justify tolerating the harms they cause. Here I will directly address only bigoted insults, because I believe they provide the clearest case for justifying *prohibition* of a category of speech. However, I hope to leave open the question whether other forms of bigoted or assaultive speech might justifiably be subject to forms of restriction proportionate and tailored to both their harmfulness and potential contributions to the free exchange of ideas.

The phrase ‘bigoted insults’ is meant to distinguish the form of speech I mean from some close relatives. Within my Millian framework, some of the speech that critical race theorists refer to as “assaultive”⁴⁷ might qualify as speech that is better expressed and answered rather than prohibited. Additionally, although bigoted insults are sometimes referred to as ‘hate speech,’ this phrase can be misleading. Those who utter bigoted insults may not hate their target or be motivated by hate; instead, they are often *indifferent* to how their victims are affected⁴⁸ by their attempts to gain acceptance to, or bond with, a social peer group by rejecting an out-group.⁴⁹ Although this sort of indifference may be as morally undesirable as hatred, its relative prevalence in society means

⁴⁴Ibid., pp. 23–26.

⁴⁵Ibid., p. 29.

⁴⁶Ibid., pp. 148–52.

⁴⁷*Words that Wound*, p. 42.

⁴⁸As Michelle Alexander notes, racial hostility is not required to maintain a racial caste system; society’s indifference to Black suffering is all that is needed. See Michelle Alexander, *The New Jim Crow: Mass Incarceration in the Age of Colorblindness* (New York: The New Press, 2012), pp. 203–04.

⁴⁹See, e.g., Larry May and Robert Strikwerda, Men in Groups: Collective Responsibility for Rape, *Hypatia* 9.2 (1994), 134–51; Michael Kimmel, *Guyland: The Perilous World Where Boys Become Men* (Harper Collins, 2008), pp. 186–87.

that bigoted insults contribute more to a bigoted social climate than we would expect if only unabashed bigots uttered them.

Richard Delgado and Jean Stefancic seem to have in mind what Fricker calls *hermeneutical injustice* when they assert that what judges hear and read about marginalized peoples largely comes from law, their own interpretive community. Law's canon consists of terms like 'just,' 'fair,' 'equal,' and 'deserving.' These concepts and meanings come from an existing paradigm, which appears to construe them as they ought to be, or even as they plainly are. Humans are psychologically inclined to believe that we deserve our advantages because they are the product of our efforts in a fair system. If this is the paradigm, and it excludes consideration of systemic oppression, then demands for change will be heard as "incoherent or unprincipled."⁵⁰

For example, institutional restrictions on bigoted insults are frequently referred to by the derogatory word 'censorship.' But censorship historically refers to a powerful government silencing weaker private speakers who are their unpopular critics or political dissidents. Those who utter bigoted insults target vulnerable people rather than criticize public figures or institutions. Restrictions on bigoted, targeted insults are not equivalent to government attempts to control the public narrative or to silence dissent. The only idea that bigoted insults express is *You do not deserve my respect or regard; you are not my social or moral equal.* This can strengthen identity-prejudicial credibility deficits and drive victims away from the discussion. Thus, hate speech regulations actually may enhance dialogue rather than hampering it. Racial and ethnic slurs "evoke and reinforce entire cultural histories of oppression and subordination." Even the most accomplished people in subordinated groups are vulnerable to the least accomplished bigot's incivility and disparagement.⁵¹ Speech "at least in the grand dialogic sense [. . .] presupposes rough equality among speakers."⁵²

The main public purpose of free speech protections, for Mill, is to safeguard a public space for opinions to be shared, for debate to take place, and for rational and reasonable people to both argue for and amend their viewpoints. Bigoted insults do not deserve protection under this rationale because they cannot reasonably be understood as opinions tendered for consideration and they are *not answerable*. We should also acknowledge that a stringently protected *marketplace of ideas* is of far greater value to dominant members of society than to oppressed members, with the latter bearing the brunt of its costs.

First, bigoted insults cannot reasonably be understood as opinions or viewpoints to be considered or discussed. Speech content can function differently in different contexts. The word 'Fire!' may serve as a warning to fellow occupants of a burning building, or it may serve as a military command to discharge weapons. The function of expressions within particular social contexts is usually not difficult to discern for people familiar with the culture and language.⁵³ The effect and sometimes the intent of bigoted insults is to "reassert and reinscribe" the speaker's social power over the person insulted as a member of a socially stigmatized group. As Delgado and Stefancic argue,

⁵⁰Richard Delgado and Jean Stefancic, *Must We Defend Nazis? Hate Speech, Pornography, and the New First Amendment* (New York: NYU Press, 2018), p. 123.

⁵¹*Ibid.*, pp. 94–95.

⁵²*Ibid.*, p. 114.

⁵³Melanie Beres, *Sexual Miscommunication? Untangling Assumptions About Sexual Communication Between Casual Sex Partners, Culture, Health & Sexuality* 12 (2010), 1–14.

Racial insults and remarks are among the most pervasive means by which discriminatory attitudes are imparted, communicating the message that distinctions of race are ones of merit, dignity, status, and personhood. Not only does the listener learn and internalize these messages, they color our institutions and are transmitted to succeeding generations.⁵⁴

Bigoted insults are a form of cultural propaganda that ordinarily bypasses rational processing. Hearing an unsubstantiated statement repeated tinges it with familiarity, making hearers more likely to believe it whether or not it is true.⁵⁵ Moreover, repeated exposure to bigoted stereotypes conditions people to have unconscious, implicit biases that are not available to reflection. These biases can lead people to errors in cognition and judgment, and those errors can lead to prudential and moral mistakes. Although a person understands the principles surrounding what is in their self-interest or what is the right thing to do, factual mistakes about their environment and the people with whom they inhabit it can cause their judgment to misfire. People do not consent to implicit biases, which interfere with their agency. Everyone in a bigoted society therefore suffers unconsented-to harm when bigoted insults, jokes, and statements invade their character and judgment and mislead them. But bigoted utterances have the most devastating impacts on those who are also oppressed by the unjust cultural hierarchies they sustain. Bigoted insults are not put forward as a point of view to be tested, contested, and defended. They are targeted at particular individuals just as a physical blow would be, to harm members of stigmatized groups. They are a form of assault and subordination, not a form of communication. Physical assaults can also express a viewpoint, or at least an attitude toward the victim, but their function is only incidentally, not primarily, expressive. Any viewpoint they express can be expressed in a less harmful, more constructive manner.

Second, viewpoints and opinions are answerable; the accuracy of their content can be challenged, buttressed, and debated in a suitable context. But this is not true of an attack by bigoted insult. Suppose a group of white college men yell a racist, sexist insult (e.g., “N___ whore!”) at a female Black student walking across campus at night. What is she to say to counter that speech? “No, actually, I never have been paid for sex, so ‘whore’ is inaccurate besides being derogatory to sex workers; and N___ is a word with a history connected to intimidation, terrorism and murder of Black people, and which, because of false beliefs about the intelligence and work ethic of Black people, has operated to deny them recognition of their equality and full personhood.” Would that prompt an apology? Would it make her attackers recognize the truth of the words spoken, so that they stop shouting bigoted remarks at fellow students?⁵⁶ The likely effect would be to place the female student in danger of physical violence, and/or invite further insults and mockery. This is not an exchange of ideas, does not contribute to anyone’s self-development, and does not bring anyone closer to truth. It does not function as communication but as an assault.

As noted in Part I, Mill recognizes free speech as having both public and private value. Admittedly, it might be just as important to the white students shouting insults

⁵⁴Delgado and Stefancic, *Must We Defend Nazis?*, pp. 7–8.

⁵⁵Roselyn J. Lee-Won et al., Source Magnification of Cyberhate: Affective and Cognitive Effects of Multiple-Source Hate Messages on Target Group Members, *Media Psychology* (2019); see also Tom Stafford, How Liars Create the Illusion of Truth, *BBC Future* (26 October 2016) <<http://www.bbc.com/future/story/20161026-how-liars-create-the-illusion-of-truth>> [accessed 1 June 2020].

⁵⁶My example is an adaptation of Delgado and Stefancic’s example. *Must We Defend Nazis?*, pp. 68–69.

that they are free to do so as it is to the African-American student to avoid those insults. Ideally this would not be true, but in speaking of subjective sources of individual fulfillment, it is difficult to deny this sometimes might be true. If these hypothetical persons' interests are equally matched, however, the tie is easily broken by the other important dimension of free speech: the public interest in truth-seeking, equal concern for constituents, and aggregate human happiness. Public values line up squarely against permitting such an utterance in this context.

Even under the First Amendment to the US Constitution, which so stringently protects freedom of speech, exceptions are made when speech causes harm – that is, sets back recognized individual interests – and thereby collides with other important values, such as reputation, security, and ownership of ideas. Publishing false statements that damage people's reputation can make one liable to defamation claims. Making illegal threats and inciting riots are unlawful. Copyright protections restrict use of others' ideas to incentivize creativity. Mill would agree that significant risks of harm to reputation, personal security, and property justify speech restrictions, *even when the restricted expression represents an opinion or viewpoint*.⁵⁷ Significant risks of harm from bigoted insults are, likewise, antithetical to important social values, such as equal opportunity and equal concern for people's comparable welfare interests.⁵⁸ Why are these types of harmful speech viewed so differently that there is general agreement that defamation, illegal threats, inciting riots, and violating copyrights should not be protected activities, while equally harmful bigoted insults receive stringent legal protection?

Bigoted insults are an important weapon to help those in power retain their privilege, whether or not they are used deliberately for this purpose. Legal rules protecting property and reputation are also designed to safeguard the interests of powerful people in society, who are far more likely to benefit from these speech restrictions and seek remedies using their frameworks. But because bigoted speech and its harmful effects are not widely recognized as part of an oppressive system, but regarded as a series of unrelated incidents – and because the harms are not recognized as such, due in part to epistemic injustice – protecting bigoted speech makes free speech absolutists appear principled, as though they are claiming the high ground.⁵⁹ Their stance appears objective because it regards itself as neutral between different viewpoints, where civil rights advocates are portrayed as desiring to suppress speech that offends overly sensitive people, catering to their subjective preferences. This distortion arises from an institutional paradigm that accepts a civil liberties framework and rejects a civil rights framework, given historical circumstances in which a dominant majority built social institutions around a paradigm sensitive to, and suited to, their own needs and interests.

The formalistic First Amendment jurisprudence that struck down Stanford's speech code would accept restrictions on speech for the alleged purpose of preventing *violence*: street fights provoked by "fighting words." That these words were intended to insult on the basis the target's sex, race, color, handicap, religion, sexual orientation, or national or ethnic origin supposedly constituted viewpoint discrimination: it showed government favoritism for one point of view over another. Targeted compliments based on the named categories (whatever that might mean) were permitted, the court

⁵⁷Piers Norris Turner, Authority, Progress, and the 'Assumption of Infallibility' in *On Liberty, Journal of the History of Philosophy* 51.1 (2013), 93–117 (p. 105).

⁵⁸Delgado and Stefancic, *Must We Defend Nazis?*, p. 153.

⁵⁹*Ibid.*, pp. 158–59.

admonished, but targeted insults were forbidden. So, according to the court, a positive view of people based on their identity within *named categories* was impermissibly favored by the state over a negative view.

The court treated all members of *named categories* as symmetrically situated: as if white people could be insulted on the basis of race equally with Black people, or heterosexual cisgender people could be insulted on the basis of sexual identity/orientation equally with LGBTQ+ people.⁶⁰ The asymmetrical vulnerability to insults of some members of these identity categories (e.g., “race” and “sexual orientation”) seems to have gone unrecognized.

While the court said *all* fighting words might constitutionally have been prohibited, Stanford prohibited only words whose insults were based on identities associated with what the court called “favored” categories (race, sex, etc. – implicitly equating these with subordinated subgroups – people of color, women, etc.). Effectively, fighting words doctrine affords protection from insults to straight white men: members of a dominant social group who take a smaller risk when they become angry and respond to insults with violence. But that doctrine does not *explicitly* invoke a demographic category, so courts accept it, even though it gives more protection from insult to the numerical minority group (straight white men) whose interests the institution of law was designed to protect. Worse, it ignores the way bigoted vilification functions not as speech – as communication – but as an assault on members of a vulnerable group. The First Amendment does not protect those beating or killing members of marginalized groups, even when doing so clearly expresses an attitude toward the victim or a group to which they belong. How much harm must be threatened to constitutionally justify restraining expressive conduct? Or does it matter more who would suffer the harm, and whose expressive conduct would be restrained? In an institutionally racist society, colorblind or neutral policies (such as the fighting words doctrine) favor the interests of dominant groups, who are seen as the neutral, ordinary people, and do not adequately protect those at the margins of the system.

It is sometimes argued that the civil rights movement depended on the freedom of activists to speak against the status quo. But in fact, a First Amendment free speech guarantee has not been much help to African Americans. Contained in a Constitution that recognized slavery, it never protected the free speech of enslaved persons. The Founders meant to protect the sort of speech that educated, upper class white men engaged in. Civil rights activists often spent time in jail for speaking out against abusive state and local governments and the white establishment.⁶¹ The First Amendment, as Delgado and Stefancic argue, “is far more valuable to the majority than to the minority, more useful for confining change than for propelling it.”⁶² The argument that *more speech* is the cure to bad speech ignores the disproportionate dangers that marginalized people face, when answering is even possible (graffiti and anonymous messages, for example, cannot be answered).

How might Mill’s impartial agent⁶³ attempt to reconcile the civil rights and civil liberties paradigms in creating a just social institution? Here is a theory I regard as

⁶⁰See Harold Blum, Racial and other Asymmetries, in *Philosophical Foundations of Discrimination Law*, ed. Deborah Hellman and Sophia Moreau (Oxford 2013), 182–200 <doi:10.1093/acprof:oso/9780199664313.003.0010>.

⁶¹*Must We Defend Nazis?*, pp. 65–66.

⁶²*Ibid.*, p. 67.

⁶³Mill, *Utilitarianism*, CW X, p. 218.

plausible, given what has been said so far. For all of the reasons Mill offers regarding the social value of speech in discovering truth and living better, an opinion or viewpoint should not be restricted, no matter how offensive, unless it is also harmful. Harmful speech that does not convey an opinion or viewpoint, however, generally will not deserve protection. Moreover, *the means* of conveying even an opinion or viewpoint can justifiably be restricted if it risks significant harm to nonconsenting others (the harm principle), and should be so restricted to the extent the restriction does not risk more harm than would occur in its absence (the Greatest Happiness Principle). Additionally, as Mill explicitly acknowledges, majorities must not maintain rules to favor their own situations simply because they wield the power to do so: all persons' like interests (for example, in free speech, personal security, health, and material means) should receive the same degree of protection. And generally, only in retrospect do we realize that society did not protect everyone's interests equally.⁶⁴ We need to be more vigilant and proactive.

Suppose in a college class, or in a debate in the public square, a white man expresses the view that men are naturally better at mathematics, or that they make better political leaders; or the view that African-American poverty is a consequence of poor work ethic. These claims articulate bigoted stereotypes and are likely to offend many people. However, as Mill recognizes, prohibiting them from being spoken does not permit the opportunity for their refutation, or for people to amend their beliefs in response to arguments and evidence. Expression of these opinions should be protected, and others should have the opportunity to refute them. Otherwise, both points of view appear equally valid, with two sides in disagreement but with no fact of the matter to be established or recognized. (And of course, it should not always fall on members of targeted groups to refute these arguments; all are obligated to take on their share.)

An opinion expressed in class is markedly different from the scenario considered earlier, in which a group of white male college students target a racial insult at a particular African-American peer as she passes by. This insult is far more like a physical assault in its intention to harm, and its effect on the woman's well-being interests, than it is like offering the opinion, in general, that Black people are inferior or that women are only valuable as sex objects. The latter express opinions that can be argued against with evidence and/or principles in a context that does not personally threaten a particular individual's security. In fact, stated as an opinion either claim would be very difficult to support. But the purpose of bigoted insults is not to express a viewpoint or opinion; it is to harm a vulnerable person and reinforce one's own dominant social status. Without institutional oppression, bigoted speech would not raise the concerns it does in a society where marginalized people are oppressed within a social hierarchy. Yet, without an oppressive social hierarchy it would be difficult to understand the motivation for, or cause of, bigoted speech in the first place. Bigoted insults do not contribute to debate or the search for truth, and should not be protected on the grounds that (even profoundly offensive) opinions and viewpoints should be protected.

Feinberg argues that in certain cases, profound offensiveness should be a moral justification for coercively prohibiting an unwilling person's exposure to that behavior. He offers complex tests and conditions for formulating policies aimed at preventing overreach of the offense principle. In my view, however, Mill had it right. Even if Feinberg's policy positions turned out to be justified, I maintain that in principle the mere fact that a majority shares certain sensibilities – in the absence of unconsented-to harm – should

⁶⁴Ibid., p. 259.

not be accepted as grounds for coercive restriction of behavior engaged in by minorities. That would simply be a form of the tyranny of the majority that Mill sagely warned against. It disproportionately – and arbitrarily – empowers whatever group happens to form a majority at a particular time to impose coercive restrictions that constrict minorities' scope of liberty. As a liberty-limiting principle, the harm principle is the only one that adequately protects minorities.⁶⁵

Moreover, if bigoted insults were officially rejected as valued and protected speech, their role in reinforcing institutional racism would be undermined, and social progress might conduce to a point where people do not receive personal fulfillment from asserting their dominance over others or reinforcing systems in which they occupy unjustly privileged positions. Mill recognized how social values can evolve over time, and how attention to values and empirical facts, together with social policies deliberately aimed at cultivating personal development and empathy for others, can change our individual preferences in ways that help us live more fulfilling lives in greater harmony.⁶⁶ Optimal early socialization might even reduce the tendency discussion now has to undermine receptivity to opposing opinions and exacerbate sectarianism. This would enhance discussion's tendency to produce knowledge. We could also do more as a society to educate children – over whom paternalism is appropriate⁶⁷ – about the harms that bigoted speech, and especially bigoted insults, can cause, and about the falsity of stereotypes that are propagated via these means. Mill explicitly acknowledges that “the love of domineering over others” is an immoral disposition that justifies moral disapprobation.⁶⁸

I conclude that it is not morally justifiable to use coercive means to attempt to silence or restrain merely offensive speech, although some harmful speech that has gone unregulated can, with moral justification, be subject to social regulation. Bigoted insults are an example of harmful speech that has been protected by society without justification. Using the Millian framework I propose, society is likely justified in restricting other forms of harmful speech that it now protects.⁶⁹

⁶⁵Mill's notion of consent might also need expansion along the lines Onora O'Neill suggests in *Between Consenting Adults*, *Philosophy & Public Affairs* 14 (1985), 252–77.

⁶⁶Mill, *The Subjection of Women*, *CW* XXI, p. 272; *Utilitarianism*, *CW* X, p. 213.

⁶⁷Mill, *On Liberty*, *CW* XVIII, p. 282.

⁶⁸*Ibid.*, p. 279.

⁶⁹I am very grateful to Nathaniel Goldberg, Christopher Macleod, Dale E. Miller, Angela Smith, and Piers Norris Turner for detailed and helpful comments and suggestions on drafts, to Julie A. Woodzicka for valuable references, and to Jim Casey, Art Goldsmith, Leah Naomi Green, Chris Handy, Howard Pickett, and Katharine L. Shester for early stage discussions that were critical for developing my argument. My thanks also go to the audience at the Philosophy, Politics, and Economics Society 2019 Meeting. This work was supported by a Lenfest Grant from W&L University. Support for the publication of this article was provided by the Class of 1956 Provost's Faculty Development Endowment, the Philosophy Department, and the Office of the Dean of the College at W&L University.