Canadian Journal of Law & Jurisprudence XXXV No.1 February 2022, 31-57
 doi: 10.1017/cjlj.2021.18

Words That Harm:
Defending the Dignity Approach to Hate Speech Regulation

Chris Bousquet
Department of Philosophy, Syracuse University, USA

Abstract

The dignity approach to racist hate speech regulation maintains that hate speech ought to be regulated because it impugns targets' dignity and poses a threat to their equal treatment. This approach faces the significant causal challenges of showing that hate speech has the power to erode its targets' dignity and that regulations can successfully protect that dignity. My aim is to show how a friend of the dignity approach can resolve these challenges. To do so, I borrow insights from the critical legal studies (CLS) approach to hate speech. Specifically, I argue that hate speech can erode its targets' dignity 1) by constituting an act of discrimination, and 2) by enacting norms that call for treating targeted groups as inferior. Yet while I maintain that the CLS approach offers valuable resources for shoring up the dignity approach, I reject the CLS approach in favor of the dignity approach.

1. Introduction

In 2017, Justice Alito delivered the opinion for *Matal v. Tam*, the U. S. Supreme Court's most recent ruling on hate speech. Rejecting a restriction on hate speech, Alito argued, "Speech that demeans on the basis of race, ethnicity, gender, religion, age, disability, or any other similar ground is hateful; but the proudest boast of our free speech jurisprudence is that we protect the freedom to express 'the thought that we hate.'"[1] This assertion by Alito that we must tolerate even "the thought that we hate" is indicative of American legal orthodoxy concerning hate speech. The idea is that the government should not regulate speech merely because many find it offensive, for this risks silencing minority views in a way that's antithetical both to political deliberation and basic freedoms. However, some have challenged this legal orthodoxy, arguing that hate speech regulations are in fact consistent with a healthy respect for free speech and promote rather than hinder guarantees of basic rights. It is my purpose in this paper to give such an argument, offering a justification of racist hate speech regulations that embraces the free speech principle as it exists in Western political thought.

In particular, I argue in favor of what I call the dignity approach to racist hate speech regulation. This approach, exemplified most prominently in the work of Jeremy Waldron, emphasizes the way that racist hate speech impugns targets' social

1. *Matal v Tam*, 137 S Ct 1744 at 1764 (2017).

standing and poses a threat to their equal treatment, thereby outweighing the value it has as speech. As articulated in existing literature, however, this approach faces the significant causal challenges of showing that hate speech can in fact erode its targets' dignity and that regulations can successfully protect that dignity. My aim is to show how a friend of the dignity approach can resolve these challenges. To do so, I borrow insights from what I broadly term the critical legal studies (CLS) approach to hate speech. Specifically, I argue that hate speech can erode its targets' dignity by: 1) constituting an act of discrimination, and 2) enacting norms that call for treating targeted groups as inferior. Yet while I maintain that the CLS approach offers valuable resources for shoring up the dignity approach, I reject the CLS approach in favor of the dignity approach because the former misunderstands the relation between hate speech and the interests underlying free expression.

2. Hate Speech

Before proceeding, I should clarify what it is I mean by racist hate speech. As a starting point, racist hate speech is speech that vilifies, demeans, or threatens members of a marginalized racial group[2] based on their membership in that group. Caleb Yong offers a useful taxonomy of four types of speech that might satisfy something like this definition: targeted vilification, diffuse vilification, political advocacy for exclusionary or eliminationist policies, and adverse assertions of fact or value.[3] By targeted vilification, Yong denotes speech whose dominant intention is to wound or intimidate, motivated by hostility towards some racial group and directed at an individual member or multiple members of that group. Diffuse vilification has the same content as targeted vilification, but is directed not specifically at members of the racial group, but a sympathetic public audience. Political advocacy for exclusionary or eliminationist policies includes public support for heinously racist policies like segregation or genocide. And adverse assertions of fact or value include general indictments of racial groups as a whole, such as expressions of stereotypes that members of some group are criminal or lazy.

The types of hate speech I have in mind in this paper will include utterances that fall into each of these categories, though not necessarily all utterances in each category. Thus while I take Yong's classification as a useful indication of the kinds of speech that may count as hate speech, I reject his argument that utterances in the first three categories are regulable, while those in the fourth are not. This is because it's possible for an utterance belonging to any of these four categories to create dignitary harms that outweigh its value as speech.

Yet I also doubt that all utterances fitting this definition ought to be regulated, for there likely will be utterances that vilify or demean members of marginalized racial groups in virtue of their race and yet do not enact dignitary harms

2. One might wonder whether racist hate speech can also apply to utterances directed at non-marginalized groups. I will not rule out the possibility here, but it is not my focus in this paper.
3. Caleb Yong, "Does Freedom of Speech Include Hate Speech?" (2011) 17:4 Res Publica 385 at 394-402.

https://doi.org/10.1017/cjlj.2021.18 Published online by Cambridge University Press

outweighing their communicative value. The purpose of this paper is not to delineate the precise extension of speech that ought to be regulated, but rather to show *why* certain kinds of speech should be regulated. That said, it's worth considering some examples to get a sense of those utterances to which my argument applies. The following are all utterances that, by my lights, constitute dignitary harms that outweigh the free speech interests they implicate. To illustrate the scope of my thesis, I've chosen examples of speech that fall within each of the four categories that Yong outlines: targeted vilification, diffuse vilification, political advocacy for exclusionary or eliminationist policies, and adverse assertions of fact or value. While I'll table until later the question of the form this regulation should take, it's worth flagging here that penalties for hate speech will certainly come in degrees, reflecting the degree of dignitary harm perpetrated by such speech.

Consider the following four examples:

> **(Racist Subway Heckler):** A white man on a subway directs racial slurs toward a Black woman standing on the other side of the car and tells her to "go back to Africa." (targeted vilification)
>
> **(Racist Subway Preacher):** Instead of directing slurs at the Black woman herself, the man directs his comments to other white individuals on the subway, yelling, "If we don't send these people back to Africa, they're going to take over our country." (diffuse vilification)
>
> **(Nazi Facebook Group):** Individuals create a public group on Facebook devoted to "carrying on and implementing the legacy of the Third Reich" in which participants debate the best political strategies for keeping Jews, immigrants, and people of color out of public office and excluding them from schools and places of business attended by white people. (political advocacy for exclusionary or eliminationist policies)
>
> **(Islamophobic Pamphleteer):** A white woman hands out pamphlets on street corners purporting to tell "the truth" about Muslim people in America, claiming they are inherently less intelligent, lazier, and more violent than white people (adverse assertions of fact or value).

These are all examples of speech that ought to be regulated on my account. Now, while I do not purport to give necessary and sufficient conditions for the kinds of speech that produce dignitary harms outweighing their value as speech, there is an important commonality between these four examples that might provide the starting point for such an account. That is, in each of these cases, a non-member directs towards members of a marginalized racial group an utterance that either explicitly or implicitly expresses a belief with four characteristics: a) it is false; b) it attributes to members of a marginalized racial group some deficiency of character; c) that deficiency is supposedly inherent to membership in the racial group; and d) a group's actual possession of this deficiency would offer a strong reason to exclude them from social and political life and/or give their interests less consideration. Each implies that simply by virtue of membership in some racial group, targets possess some severe moral or intellectual deficiency that

https://doi.org/10.1017/cjlj.2021.18 Published online by Cambridge University Press

would ground their exclusion from full participation or concern, maintaining that they are too unintelligent or dangerous or unethical to be trusted or cared about. Such utterances are extreme expressions of racial discrimination, which on Tommie Shelby's view "is operating when a so-called racial characteristic (or set of characteristics) possessed by or attributed to the members of a social group is wrongly treated as a source or sign of disvalue, incompetence, or inferiority."[4] Even the mere use of a slur, for instance, communicates a message in favor of exclusion, drawing on a history of racial stereotypes to mark targets as inferior and unworthy of equal participation and consideration. Indeed, as I will outline in more detail later, it's precisely this message that makes such utterances so pernicious.

I do not, it's worth noting, intend this to be an account of the extension of racist hate speech. It would be consistent with my view if the kinds of utterances I've just described comprise only a subset of racist hate speech, and therefore that it is only a subset that ought to be regulated. That said, for the sake of simplicity, in the remainder of the paper I will use 'racist hate speech' to denote only the kinds of utterances I have just outlined, but remain agnostic as to whether such utterances truly exhaust the category. In the following sections, I will argue the harms that such speech produces by excluding targets from the full privileges of citizenship and denigrating their status as equal members of society outweigh the free speech interests they promote.

3. The Free Speech Principle

In this paper, I take it as given that freedom of expression is valuable and that speech is therefore deserving of special political protections. I accept the traditional free speech principle (FSP) as characterized by Yong: "a *distinct* principle which *goes beyond* a general principle of negative liberty ... requiring a more stringent standard of justification for the restriction of speech than for other activities."[5] I take this principle as a plausible assumption shared by most interlocutors in the hate speech debate and the foundation of the political right to free speech.

As I see it, there are three key justifications for this free speech principle, rooted in three interests. The first and probably most commonly cited is based on what Joshua Cohen calls the deliberative interest.[6] This justification focuses on the proclivity of free expression to promote the acquisition of truth, enabling people to obtain information that aids them in navigating the world successfully and pursuing the good. The second is rooted in what Cohen calls the expressive interest, the direct interest people have in articulating their beliefs and attitudes on

4. Tommie Shelby, "Justice, Deviance, and the Dark Ghetto" (2007) 35:2 Philosophy & Public Affairs 126 at 131.

5. Yong, *supra* note 3 at 387 [emphasis in original]. Yong credits this principle to Frederick Schauer, *Free Speech: A Philosophical Inquiry* (Cambridge University Press, 1982); Kent Greenawalt, "Free Speech Justifications" (1989) 89:1 Colum L Rev 119; TM Scanlon, "A Theory of Freedom of Expression" (1972) 1:2 Philosophy & Public Affairs 204.

6. See Joshua Cohen, "Freedom of Expression" (1993) 22:3 Philosophy & Public Affairs 207.

https://doi.org/10.1017/cjlj.2021.18 Published online by Cambridge University Press

what is valuable. The third focuses on the nature of democratic decision-making and the interest in ensuring that all citizens are able to express their views and influence political decisions.[7]

Given the FSP, there are essentially two ways of arguing that some speech is regulable. The first is to show that the relevant speech is uncovered, meaning it is unrelated to the interests that underlie free speech protections and therefore outside the scope of the right to free speech.[8] As Yong puts it, uncovered utterances are "so unrelated to any conceivable justification of free speech that they do not arouse even the minimal concern of the FSP."[9] The other is to show that the relevant expression, while covered, is nonetheless unprotected because other considerations outweigh the FSP. My argument for racist hate speech regulation will take the latter approach, and following my positive account, I will show why the former is unsuccessful.

Yong also draws a distinction between speech that is regulable and speech that ought to be regulated. Speech is regulable when it is uncovered or unprotected by the FSP, but according to Yong, meeting these criteria does not yet establish whether it ought to be regulated. As he points out, there may be reasons independent of free speech concerns that ultimately count against regulating some speech.[10] I think, however, that this division applies only to uncovered, not unprotected speech. It certainly may be the case that even though speech fails to advance free speech interests, it still shouldn't be regulated because the speech serves other interests or because regulation would cause some harm unrelated to free expression. Yet if speech is unprotected, it has already been established that the reasons in favor of regulation outweigh those against, and by a good margin. For speech will only be unprotected if the balance of other reasons outweigh the FSP. This means not only that the weight of independent considerations ultimately count in favor of regulation, but that the magnitude of these reasons in favor of regulation is enough to outweigh the free speech interests advanced. Because my argument is that racist hate speech is unprotected, it serves both as an argument that such speech is regulable and that it ought to be regulated.

4. The Dignity Approach

I will here argue for a form of what I call the dignity approach to racist hate speech regulation, reflected most prominently in the work of Jeremy Waldron.[11] This approach does not seek to show that racist hate speech is uncovered by our free speech principle or that it does not advance any free speech interests, but rather maintains that the harms done by racist hate speech outweigh its

7. See generally, Cass Sunstein, *Democracy and the Problem of Free Speech* (The Free Press, 1993).
8. The distinction between covered and uncovered speech was originally developed in Schauer, *supra* note 5.
9. Yong, *supra* note 3 at 388.
10. *Ibid*.
11. See Jeremy Waldron, *The Harm in Hate Speech* (Harvard University Press, 2014).

https://doi.org/10.1017/cjlj.2021.18 Published online by Cambridge University Press

proclivity to promote these interests. Specifically, Waldron argues that racist hate speech impugns the dignity of its targets, where dignity is understood not as some subjective sense of personal worth, but rather as one's objective social standing. Such speech threatens what Stephen Darwall calls "recognition respect": individuals' entitlement "to have other persons take seriously and weigh appropriately the fact that they are persons in deliberating about what to do."[12] Hate speech labels its targets as inferior and unworthy of the rights and privileges of full citizens, undermining their place as equal, respected members of society. Given the still fragile state of racial equality, Waldron concludes that the persistence of hate speech is incompatible with the liberal duty to preserve the equal dignity of all citizens. This is the obligation to ensure that all citizens retain equal consideration in decision-making, equal recognition as human agents capable of making and executing plans for their lives, and equal opportunity to participate in the cultural, professional, and political spheres. Hate speech presents a kind of environmental hazard that makes the fulfillment of this duty more difficult, polluting the social environment with the message that members of certain groups ought not have equal opportunity, thereby perpetuating racial hierarchies.

Waldron draws an analogy between racist hate speech regulation and laws against defamation. According to Waldron, such laws, which prohibit the publication of materials that attack the character of others via false accusations, are intended to protect the social and legal status of all members of society. Defamation can destroy its targets' reputation, driving others to deny them equal consideration and participation. Publishing false material claiming your neighbor is a pedophile, for instance, will likely inspire others in the neighborhood to seek to drive him out, employers to deny him job opportunities, and generally others in society to give his interests less weight. In short, such defamation unjustifiably strips him of his equal status in social and political life. One can justify hate speech regulations on the same grounds: by falsely marking members of certain racial groups as inferior, dangerous, violent, and ultimately unworthy of equal respect or social standing, hate speech strips its targets of their dignity, calling for others to deny them equal recognition. The U.S. Supreme Court used something like this justification for hate speech regulation in *Beauharnais v. Illinois* in 1952, one of the first and only cases to uphold hate speech laws.[13] The Court upheld an Illinois statute prohibiting speech degrading to Black Americans under a group libel principle. Justice Frankfurter argued that by calling on others to deny Blacks equal status, hateful and defamatory messages inhibit peaceful coexistence in a diverse society. The crucial consideration was the tendency of such expression to erode the dignity of Black citizens and preclude the successful operation of a society of free and equal members.

It is because of the still-fragile status of many non-dominant ethnic groups in the United States that the kind of speech that I have singled out—speech expressing extreme discrimination against members of marginalized racial groups—is a

12. See Stephen Darwall, "Two Kinds of Respect" (1977) 88:1 Ethics 36 at 38.
13. *Beauharnais v Illinois*, 343 US 250 (1952).

https://doi.org/10.1017/cjlj.2021.18 Published online by Cambridge University Press

particularly salient threat to maintaining equal dignity. As Waldron notes, even formal racial equality is a recent accomplishment in the United States,[14] only instituted within the last fifty years or so and arguably still not fully realized. Moreover, de facto racial discrimination certainly persists, both in interactions among members of society and between individuals and the government. As a result, the dignity of members of marginalized races is still tenuous, not something to be taken for granted. There are still many inclined to believe that members of these groups are inferior and unworthy of the privileges of Americans, and willing to deny them equal treatment. As a result, expressions of extreme discrimination against members of these groups pose a particularly grave threat to their dignity. Hate speech directed at members of marginalized racial groups confronts dignity already under threat, and therefore risks eroding that dignity to a paltry state. It is more likely to communicate to members of those groups that they are not equal members of society because they are confronted with this message in so many other parts of life. It is also more likely to exclude, to lower the estimation of group members in the eyes of others, and to perpetuate norms of discrimination. This threat justifies intervention: the dignity of the members of these groups must be proactively protected and supported. There may come a day when the status of members of marginalized groups is well-established enough that hate speech laws won't be necessary, but we are far from that moment.

Waldron's thesis is that the harm done by stripping away its victims' dignity outweighs any free speech interests underlying hate speech. He thinks an honest defense of hate speech regulation must acknowledge the costs of such regulation, i.e. the free speech interests that are suppressed.[15] However, he is not explicit on how this weighing of interests is supposed to go. He does not explain whether his conclusion is based on a consequentialist calculation that pits the utility of free expression against the harms of racist hate speech, a weighing of rights against other rights, or some hybrid of the two. As Brian Leiter notes, the result is that it is not entirely clear why the dignitary harms of hate speech warrant regulation, while other harms to vulnerable people do not.[16]

In my estimation, the best version of the dignity approach focuses on a weighing of two conflicting rights: the right to equal status and the right to free expression. This right to equal status expresses the entitlement of members of liberal societies to be granted equal concern in political decisions, equal access to the privileges of citizenship, and equal respect from fellow citizens as valuable and autonomous individuals. This is the right to which Kenneth Karst refers when he explains that, under the equal protection clause of the 14th Amendment to the U.S. Constitution, "every individual is presumptively entitled to be treated by the organized society as a respected, responsible, and participating member."[17] There

14. Waldron, *supra* note 11 at 31.
15. *Ibid* at 147.
16. Brian Leiter, "The Harm in Hate Speech" (2012) Notre Dame Philosophical Reviews, online: https://ndpr.nd.edu/reviews/the-harm-in-hate-speech/.
17. Kenneth Karst, "Citizenship, Race, and Marginality" (1998) 30:1 Wm & Mary L Rev 1 at 1.

https://doi.org/10.1017/cjlj.2021.18 Published online by Cambridge University Press

are certain formal aspects of this entitlement: equal rights to vote and hold office, laws prohibiting institutions from discriminating on the basis of race, and decision-making procedures that consider the interests of all citizens equally. Yet Karst maintains there is also an informal element that cannot be enshrined in any single law, a mandate to create an environment that upholds the equal status of all and in which citizens accord each other recognition respect. Achieving this mandate is imperative to creating a social and political culture in which members relate to each other as moral equals, arguably the foundational feature of liberal democracy and an assurance that some identify as a precondition for legitimacy.[18] Moreover, threats on this informal aspect of equal status also endanger the formal guarantees of equality. When individuals are not treated as equals in social and political life, their other rights are under threat: lurking in the public consciousness is a sense that they can be physically abused because their pain is less important, that they can be kept from voting because their political contributions are less valuable, and that they can be silenced because their speech is unworthy of protection.

When it comes to racist hate speech, this right of targets outweighs hate speakers' right to free expression. When a white man racially abuses a Muslim woman on the subway, he sends the message that she is unworthy of the rights of other citizens, calls for others to deny her those rights, and thereby puts those rights in danger. Of course, in most liberal democracies there are already laws in place that punish violations of these other rights, but it would seem foolish to punish such violations and yet leave untouched the actions that create the conditions for attacks on person, speech, and property. The target of hate speech on the subway might suffer in silence out of a reasonable fear of incurring violence, and it would offer little consolation to her that such violence would be punished.

It should not, however, be thought that the value of the right to equal status is merely derivative of other rights. Rather, all citizens have a right to the assurance that they are equal members of the social and political world, a right not to fear that their status is up for grabs, and therefore a right not to be subjected to racist abuse. As a result, when we consider hate speech regulation, we must weigh the many rights of targets against the speaker's right to free expression. And, we must note that targets' rights are threatened more comprehensively, while the speaker's right to free expression is restricted only in one circumscribed area, i.e. in the expression of racist language that impugns targets' dignity. Targets' rights may be restricted less directly than speakers', not limited by law but threatened by other citizens, but this doesn't mean the threats are any less effective.

18. See generally, Joshua Cohen, "Deliberation and Democratic Legitimacy" in James Bohman & William Rehg, eds, *Deliberative Democracy* (MIT Press, 1997) 67; Jurgen Habermas, *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*, translated by William Rehg, (MIT Press, 1996); Niko Kolodny, "Rule Over None I: What Justifies Democracy" (2014) 42:3 Philosophy & Public Affairs 195; Niko Kolodny, "Rule Over None II: Social Equality and the Justification of Democracy" (2014) 42:4 Philosophy & Public Affairs 287; Ronald Dworkin, *Sovereign Virtue* (Harvard University Press, 2000).

https://doi.org/10.1017/cjlj.2021.18 Published online by Cambridge University Press

It should be noted that this weighing of rights and interests may not fall in favor of hate speech regulation in each and every case. If a jurisdiction tailors legislation to prohibit only the kinds of utterances I specified in Section 2, i.e. those expressing the false belief that members of some marginalized racial group inherently possess a deficiency justifying their exclusion, I think it will do quite well in targeting only those utterances with harms outweighing their value. For, the dignitary harms of such utterances are particularly egregious.[19] And yet, one might conjure up cases in which the dignitary harms of such utterances don't in fact outweigh the interests they promote, for instance in which someone must utter a slur in order to defuse a bomb. The claim of the dignity approach is merely that in many cases the dignitary harms of racist hate speech outweigh the underlying free speech interests—enough cases to justify regulation. Part of the challenge of legislation in any area is that regulation is by nature general while the details of specific cases may confound expectations. The goal ought to be to craft legislation tailored as precisely as possible to cover cases where the dignitary harms of hate speech outweigh the interests such speech promotes, and focusing on the kinds of utterances I've identified will facilitate this goal.

5. Deficiencies of the Dignity Approach

The dignity approach faces two central challenges, each outlined by Robert Mark Simpson in his criticism of Waldron's treatment of hate speech.[20] On the one hand, Simpson argues that Waldron offers insufficient evidence that hate speech can perpetuate racial hierarchies, and on the other he challenges whether hate speech regulations target the main source of harm.

a. Can hate speech deprive one of dignity?

The first critique articulated by Simpson questions the degree to which hate speech has the power to truly deprive its targets of dignity and perpetuate social hierarchy. Waldron, Simpson argues, has an obligation to show that hate speech "diminishes [targets'] status in some discernible sense, beyond merely being a manifestation of the speaker's view of them as inferior or second-class people."[21] Waldron must show that hate speech does not merely reflect the speaker's endorsement of racial hierarchy, but actively perpetuates this hierarchy, for instance by driving others to treat members of marginalized racial groups with less respect. By Simpson's lights, Waldron fails to do so. It is not clear how the utterance of a slur, for instance, would convince others to view its targets as inferior or accord them lower social and political status. This is especially true in a

19. It should not be thought that this claim is a tautology, for there is a gap between identifying speech as expressing the belief that individuals ought not be accorded equal status and identifying speech as contributing to that erosion of status. It is my purpose in the remainder of this paper to bridge that gap (i.e., to show how speech that does the former also does the latter).
20. See Robert Mark Simpson, "Dignity, Harm, and Hate Speech" (2013) 32:6 Law & Phil 701.
21. *Ibid* at 710.

https://doi.org/10.1017/cjlj.2021.18 Published online by Cambridge University Press

community in which there's already a formal and informal endorsement of the ideals of racial equality and dignity, as is the case in most Western countries, including the U.S. Given that our laws guarantee equal respect and overwhelming majorities agree that people shouldn't be treated differently based on their race, hate speech seems more likely to garner derision and disgust than uptake. There is, according to Simpson, a missing causal claim in Waldron's argument regarding the ways that hate speech motivates continued prejudice. To justify his thesis that concerns over dignity outweigh free speech interests in the case of hate speech, Waldron must offer a persuasive account linking hate speech and racial hierarchy.

b. Do hate speech regulations target the right thing?

It might be thought that the primary deleterious effect of racist hate speech is not to spread racism within a society, but rather to harm what Simpson calls victims' "felt sense of security."[22] On this interpretation, hate speech does not necessarily motivate others to treat members of marginalized racial groups as inferior and unworthy of respect, but rather reveals to victims that there are those who already view and will treat them in this way. It does not so much harm targets' dignity as show them that their dignity is already under threat. Even if this is the case, however, Simpson doubts that hate speech regulation will do much good. For here it is prejudiced attitudes, not hate speech itself, which seem to be the problem. It is people's hearts and minds that deny members of marginalized races their dignity, and hate speech merely reveals this harsh reality to its targets. Indeed, if there are those in society who possess such views, it might actually be better if members of marginalized groups hear their expression. For seeing or hearing this hate at least alerts targets to those who wish them harm and enables opponents of racism to work to root out such views.

6. Strengthening the Dignity Approach

These arguments from Simpson are both compelling objections to the dignity approach. However, I do not think they are fatal. In this section, I show how proponents of the dignity approach can shore up their argument against Simpson's challenges. I argue that by borrowing a couple of crucial concepts from another approach to racist hate speech—the CLS approach—we can resolve the weaknesses of the dignity approach, establishing that racist hate speech strips its targets of dignity and perpetuates racist treatment. Specifically, proponents of the CLS approach offer two key ideas that support this claim: the idea that racist hate speech constitutes harm through discrimination, and the idea that it enacts discriminatory norms that others then follow. While critical legal theorists employ these analyses to show that hate speech, if uncovered, ought to be regulated, I argue that by supporting the dignity approach, these ideas support regulation even if hate speech is covered. In other words, I grant that hate speech

22. *Ibid* at 724.

https://doi.org/10.1017/cjlj.2021.18 Published online by Cambridge University Press

is connected to the interests underlying the FSP and deserving of the higher scrutiny that accompanies such classification. However, by impugning its targets' dignity via the mechanisms described below, the harms of hate speech outweigh the interests it advances, making it unprotected and a good candidate for regulation.

a. Racist hate speech as constituting harm

Throughout the hate speech literature, proponents of hate speech regulation frequently emphasize that hate speech constitutes rather than merely causes harm. What they mean by this is that it is not merely how others react to hate speech that does harm, but that hate speech by its very utterance does harm. To use a term from philosophy of language, harm is part of the illocutionary force of hate speech, not merely a perlocutionary effect. One of the more persuasive versions of this argument comes from Charles Lawrence, who argues that hate speech is an act of discrimination that excludes members of targeted groups by its very utterance.[23]

There are, I think, two ways that racist hate speech functions as discrimination on Lawrence's view. First, hate speech discriminates directly in that it marks its targets as second-class citizens and thereby contributes to their being exactly that. Lawrence likens hate speech to segregation, which on his interpretation of *Brown v. Board of Education*[24] was ruled "unconstitutional primarily because of the *message* [it] conveys—the message that Black children are an untouchable caste, unfit to be educated with white children."[25] Even if you don't agree with his reading of *Brown*, Lawrence here hits on something peculiar about utterances regarding fellow citizens' social and political status: they are at least partly self-validating. If a white man calls a group of Black people a racial slur and tells them they are unworthy of participating in social and political life, he contributes to the diminishment of Black people's social and political status. These utterances do not merely tell their targets that they're unequal, but actually partly constitute this inequality, since inequality is partially grounded in the stated attitudes and conduct of members of society. Of course, such utterances can't guarantee that targets don't participate in certain formal parts of political life like voting, but they can erode those informal aspects of socio-political status that Karst describes, undermining their treatment as respected members of society. As an analogy, my guarantee to my class that all will be treated as equally valued participants would mean little if I allowed members of the class to denigrate others because of their membership in a racial group. Such treatment would directly undermine this equality.

Racist hate speech therefore creates a kind of hole in the fabric of equal dignity ostensibly ensured by liberal societies. When someone utters hate speech, they

23. Charles Lawrence III, "If He Hollers Let Him Go" in Mari Matsuda, ed, *Words That Wound: Critical Race Theory, Assaultive Speech, and the First Amendment* (Routledge, 1993) 53.
24. *Brown v Board of Education of Topeka*, 347 US 483 (1954) [*Brown*].
25. Lawrence, *supra* note 23 at 59 [emphasis in original].

https://doi.org/10.1017/cjlj.2021.18 Published online by Cambridge University Press

make it the case that their target is denied equal status and recognition respect. Intrinsic to the speech act is treatment of its target as an inferior, unworthy of same the status as other members of society. In this way, each utterance of hate speech chips away at the dignity of members of targeted groups.

One might object that a single speaker has limited power to erode their target's dignity, presenting a version of what Ishani Maitra calls the Authority Objection.[26] Average hate speakers lack both formal authority to decide on individuals' status and informal authority over others' treatment of marginalized group members, and so they can only make it the case that they themselves deny a target equal status. This action, it might seem, has a minimal effect on the status of targets, as all others in society may continue to accord them equal dignity.

There are two ways of responding to this challenge: the first is to show that everyday hate speakers in fact possess some kind of authority relevant to denigrating targets' dignity, while the second is to deny that hate speakers require authority to significantly diminish targets' dignity. I think there is a good case to be made for both of these positions, and so I will introduce them in turn. I will here focus on the Authority Objection as it pertains to constituting discrimination by denigrating targets' dignity, but much of what I say here will apply *mutatis mutandis* to my argument in the next section regarding the enactment of discriminatory norms.

The first response addresses the question of what grounds a speaker's authority to denigrate individuals' status. It is clear that some have this authority: Ishani Maitra offers the example of a legislator in apartheid South Africa asserting (under the right legislative conditions) "Blacks are no longer permitted to vote," thereby denying Black South Africans social and political equality.[27] Most kinds of hate speech, however, are not uttered by people occupying formal positions of authority. And yet, holding a formal position is not the only way to possess authority. Maitra introduces two other ways that a speaker might have authority: via derived authority and via licensing.

Someone obtains derived authority when a possessor of positional authority confers it upon them. Crucially, this transfer of authority need not be explicit, but may occur by omission on the part of the initial authority. Maitra offers the case of the Bossy Student: a teacher wants her students to complete a project that requires each performing a different task, but before she can divide up the tasks, one student, Arlo, takes it upon himself to assign tasks himself in full view of the teacher.[28] The teacher doesn't intervene and thereby confers authority upon Arlo.

Similarly, while most hate speakers do not have formal positional authority, they may possess derived authority. When it comes to determining individuals' dignity, one relevant authority is the government, which decides the formal elements of individuals' social and political status by conferring their basic and

26. Ishani Maitra, "Subordinating Speech" in Ishani Maitra & Mary Kate McGowan, eds, *Speech & Harm: Controversies Over Free Speech* (Oxford University Press, 2012) ch 4.
27. *Ibid* at 94.
28. *Ibid* at 105.

https://doi.org/10.1017/cjlj.2021.18 Published online by Cambridge University Press

political rights and deciding how to weigh their interests. In virtue of failing to prohibit or sanction hate speech, the government plausibly gives hate speakers de facto derived authority. No one with formal authority challenges hate speakers' assaults upon dignity, and the government thereby cedes the authority to decide on individuals' social position.[29] In neither the Bossy Student nor the hate speech cases must the formal authority endorse the actions of the derived authority—the teacher may roll her eyes and the government may quietly deplore hate speakers' actions—but by failing to interrupt the derived authority's actions, they cede control.

A speaker can also obtain authority via licensing, which differs from derived authority in that it does not require the complicity of any formal authority. In cases of licensing, individuals implicitly grant someone authority over themselves. Maitra presents the case of the Hike Organizer as an example of licensing: a group of friends is planning a hiking trip, but no one expresses strong preferences about the details.[30] As a result, the group is taking so long to iron out the logistics together that it seems like they will never finish organizing. One of the members, Andy, decides to take the organizing into his own hands, and none of the others intervene when he presents his plan and doles out responsibilities. In virtue of failing to express any objections, the other members of the group surrender to Andy their power to decide on the trip.

In the case of hate speech, hearers can license a hate speaker with their silence. Individuals' social and political status, as I've argued, is not merely a function of the government's actions, but also of the words and actions of other members of society. As a result, just as the silence of the other members of the hiking trip grants Andy authority to determine the details of the trip, the silence of onlookers can grant hate speakers the authority to determine targets' social and political status. Their silence concedes to the position the hate speaker assigns the target, showing that they will not reject the social ranking proposed by the speaker. In effect then, audience members turn over their say in the target's status to the hate speaker.[31] And again, hearers need not agree with the hate speaker's sentiments in order to confer authority. Just as other participants in the hiking trip might quietly object to many of the details and still turn over their power to decide, audience members might disagree with the speaker and still turn over their say in the target's status.

Both of these types of authority arise in cases in which speakers lack authority in virtue of their social position. However, it is also reasonable to think that in many cases of hate speech, speakers in fact possess a kind of informal positional authority. Given the dominant social position of mostly white hate speakers, they possess a kind of de facto authority over the position of members of marginalized

29. Maitra makes a similar suggestion (*ibid* at 110), but does not identify the government as the relevant formal authority.

30. *Ibid* at 106.

31. Licensing authority over targets' social status shares much in common with Maitra's example of licensing authority to mark individuals as inferior (*ibid* at 115).

https://doi.org/10.1017/cjlj.2021.18 Published online by Cambridge University Press

racial groups.[32] Operating in the background is a structure of norms that confers on the average white person the power to at least partly determine the dignity of non-white members of society. At the very least, their opinions about targets' social status seem to carry more weight than those of the average non-white person.

I have so far argued that hate speakers possess a kind of authority that confers upon their utterances significant influence over targets' dignity. However, I also doubt that hate speakers require such authority in order to significantly affect the status of members of marginalized racial groups. Let us grant, for the sake of argument, that hate speakers lack derived authority from the government and licensed authority from others to decide individuals' status. It is still not the case that the function of hate speech is merely that one person denies members of the targeted group their equal status. First, a single utterance of hate speech interacts with targets' knowledge of the discrimination that persists in the rest of society. In a 2019 survey, three-quarters of Black and Asian Americans, as well as 58 percent of Latin Americans, said they have experienced discrimination or have been treated unfairly because of their race or ethnicity at least from time to time.[33] As a result, targets' understanding of hate speech is not that one person denies them equal status; rather, it reinforces their sense that many in society think them inferior and unworthy and will treat them accordingly. It provides further evidence that they are second-class citizens, erasing any doubts they might have had about how others will view and treat them. It might encourage targets to re-interpret questionable interactions they've had in the past, confirming their worst suspicions about the racist nature of those encounters.

One might, however, respond that this argument only establishes that single instances of hate speech can make targets think or understand that they lack equal status in the rest of society, not actually constitute this denigration of their dignity. The line between making individuals understand they lack equal status and denying them equal status is not entirely clear, but for the sake of argument we might even grant that taken in isolation, a single utterance of hate speech itself has a limited effect on the dignity of marginalized groups. However, taken together, utterances of hate speech in a society comprehensively endanger the dignity of their targets. The same 2019 survey suggests that exposure to racist language is not a marginal experience for members of marginalized racial groups, but a regular feature of their lives. 61 percent of Asians, 52 percent of Blacks, and 46 percent of Latin Americans in the U.S. say they have been subjected to slurs or jokes because of their race. Moreover, 65 percent of Americans say it has become more common for people to express racist or racially insensitive views since former U.S. President Trump was elected. This sentiment is particularly strong among Black and Latin Americans, around 75 percent of whom say

32. I thank Luvell Anderson for bringing this point to my attention.

33. Juliana Menasce Horowitz, Anna Brown & Kiana Cox, "Race in America 2019" (2019) Pew Research Center, online: https://www.pewresearch.org/social-trends/2019/04/09/race-in-america-2019.

https://doi.org/10.1017/cjlj.2021.18 Published online by Cambridge University Press

the expression of racist views has become more common.[34] Given the predominance of racist expression, we have good reason to think that the aggregate effect of hate speech on targets' status is significant: it is no longer one individual who denies members of a marginalized racial group their equal status, but many. This treatment clearly threatens the place of targeted individuals in society, as they are no longer truly accorded equal status by others. It would seem foolish to withhold regulation of hate speech because individual instances have a limited effect on dignity, while taken together they have a significant effect. It is this aggregate effect on targets' dignity that regulation seeks to stem.

Finally, as I'll detail in the next section, individual utterances of hate speech can spur the expression of additional derogatory utterances, compounding illocutionary harms with damaging perlocutionary effects.

Consider an analogy to racial discrimination by a restaurant. Arguably, if a single restaurant doesn't allow in Black folks, this has a minimal effect on the place of Black people in society. It's only one restaurant, after all, and Black residents are free to go to any other restaurant in the area. And yet, we certainly wouldn't take this to be a reason not to prohibit such racial discrimination. While any individual instance of discrimination might not have a great effect on the place of marginalized groups in society, the sum of such actions certainly does. Moreover, individual instances of racial discrimination tend to embolden others to follow suit.

So much for the Authority Objection. I have so far explained one way that hate speech constitutes discrimination for Lawrence, chipping away at targets' status by treating them as second-class citizens. The other way that hate speech can be discriminatory is that it functions as exclusionary conduct. When someone utters racist hate speech in a place of work, a school, or a public space, they ensure that the targets of their speech are unwelcome in those spaces. Their words may not literally prevent Black, Latin American, Muslim, or Asian people from entering, but they have a chilling effect, signaling that members of these groups will be treated with hostility and even violence if they enter. In so doing, they deny members of marginalized racial groups equal access to employment, education, and public spaces. This reasoning has found expression in local public accommodations laws, which prohibit hate speech in public spaces in order to ensure that people of all races truly enjoy equal access.

If hate speech does function in these ways as an act of discrimination, it clearly contributes to racial hierarchy. For understood through this framework, hate speech denies members of targeted groups their entitlement to be treated as respected members of society and equal access to some of the most important public goods enjoyed by citizens. It denies to targets the full privileges of citizenship and thereby impugns their dignity. This function of hate speech also shows that Simpson's understanding of how hate speech can contribute to hierarchy is unjustifiably narrow. It's not merely that hate speech persuades others to

34. *Ibid*.

https://doi.org/10.1017/cjlj.2021.18 Published online by Cambridge University Press

be racist, but that by its very utterance it constitutes and guarantees inequality. Thus, with respect to Simpson's mandate that we show how hate speech causes racial hierarchies, we can do him one better: it doesn't merely cause such hierarchies but manifests them.

These functions of racist hate speech ground Waldron's claim that the harms of such speech outweigh the free speech interests it promotes. Given that such speech diminishes targets' status and excludes them from certain privileges of citizenship, the harms outlined in Section 4 are manifested. The threats to targets' social and political equality and other basic rights outweigh the free speech interests implicated by hate speakers' utterances.

b. Enacting discriminatory norms

The other way that racist hate speech perpetuates racial hierarchies is by enacting racist norms that others then follow to harm targeted groups. I argue by enacting such norms, hate speech constitutes the harm of discrimination, and by galvanizing others to follow these norms, it also causes harm by encouraging racist treatment. Harm therefore resides as part of both the illocutionary force and the perlocutionary effect of hate speech.

To make my case, I will rely on Mary Kate McGowan's compelling account of how utterances of racist hate speech function as 'covert exercitives,' a more general parallel of the phenomena of conversational exercitives.[35] Conversational exercitives are speech acts that enact norms in particular conversations by changing the conversational score, a concept borrowed from David Lewis.[36] The score is a record of all conversational contributions and affects the moves participants are licensed to make. So, for instance, if I bring up the book *On Liberty* in the course of a conversation, this licenses the participants to identify *On Liberty* as the referent of the demonstrative 'the book' until a new book becomes salient. Crucially for hate speech, McGowan argues that this phenomenon generalizes to other-norm governed activities, like social interactions or workplace behavior, meaning that any move in these activities enacts certain norms by changing the score. Hate speech in particular has as its primary function enacting discriminatory norms in social situations that others then follow in order to harm marginalized groups. For instance, if a white person yells a racial slur on a public bus, they enact norms licensing participants to make targets feel unwelcome or unsafe, which if followed, do obvious harm to people in targeted groups. By enacting norms calling for racist discrimination, such utterances clearly perpetuate racial hierarchies.

Two features of covert exercitives make this categorization particularly appropriate for hate speech. For one, uttering a covert exercitive and thereby

35. See Mary Kate McGowan, *Just Words: On Speech and Hidden Harm* (Oxford University Press, 2019) at ch 4.

36. See David Lewis, "Scorekeeping in a Language Game" in *Philosophical Papers Volume I* (Oxford University Press, 1983) 233.

https://doi.org/10.1017/cjlj.2021.18 Published online by Cambridge University Press

successfully enacting a norm does not require the conscious intent to do so. When one refers to *On Liberty*, one does not usually intend to raise it to salience, at least in the sense that this is not one's conscious purpose for uttering the name. As McGowan puts it, such intentions are at the very least on the low end of our manifestness spectrum, meaning they are less conscious and less explicit than for instance my intention to refer to *On Liberty*.[37] The same goes for hate speech: one can enact discriminatory norms without consciously intending to do so, and indeed many who utter racial insults claim that they are merely "speaking their mind" or "telling it like it is."

Secondly, one does not require any formal authority to successfully enact a norm via a covert exercitive. Mirroring my response to the Authority Objection in the last section, there are two ways of parsing this: either that speakers possess a kind of derived or licensed authority to enact norms in conversational contexts, or that they do not require any authority at all to do so. On the former interpretation, the phenomena of derived authority, licensing, and unequal power relations from the last section apply equally well to enacting norms. Speakers may gain authority to enact norms in conversational contexts due to the silence of a formal authority or other conversational participants,[38] or because of their disproportionate power. Or, one might think that we do not require any authority to enact norms in conversational settings, given how ubiquitous and easy it is to enact norms in conversation. One does not seem to need any authority to make *On Liberty* salient after referring to it in a sentence, and likewise one does not need it to enact discriminatory norms via hate speech. We enact norms in conversational settings all the time without realizing it and without possessing any special authority to do so. Racist hate speech is just another instance of this familiar phenomenon, albeit a particularly harmful instance, since the norms it enacts are norms of discrimination.

Whichever interpretation one prefers, the important thing to note is that speakers need not occupy a position of power to enact a norm in a conversational setting. These two characteristics distinguish covert exercitives from standard exercitives: speech acts uttered by those in some formal authority position that enact norms, such as a demand from an executive that all employees wear collared shirts. Such utterances not only must be spoken by an authority figure in order to enact the relevant norms, but also generally require a manifest intention on the part of that figure. Hate speech is not like this. To enact a discriminatory norm and cause others to follow that norm, one need not intend to do so nor possess any particular authority. Just as when one raises some object to salience, by uttering hate speech one adds a norm to the conversational score.

McGowan argues that by acting as covert exercitives, hate speech constitutes harm. However, she defines 'constituting harm' in a strange way, as causing harm

37. McGowan, *supra* note 35 at 51.

38. This is more or less Maitra's interpretation of how hate speakers constitute harm: that they have derived and/or licensed authority to enact discriminatory norms. See Maitra, *supra* note 26.

https://doi.org/10.1017/cjlj.2021.18 Published online by Cambridge University Press

by enacting norms that prescribe harm.[39] On first glance, this might appear to be merely an effort to define away the problems that arise when one calls for regulating speech that causes harm, particularly the worry that such regulations might apply to anything that produces some subjective offense. As McGowan acknowledges, "as many who are reluctant to embrace further speech regulations are prone to point out, just about anything can be offensive to someone or other."[40]

Yet I think there's more to her insistence that hate speech *constitutes* harm than a mere definitional trick. As I said, the harm hate speech produces by enacting norms resides in both its illocution and perlocution, and my interpretation of McGowan locates the harm of hate speech in these two different places. First, hate speech directly harms its targets purely in virtue of creating discriminatory norms, because these norms themselves compromise the equal status of targets and exclude them from public spaces, workplaces, educational institutions, and beyond. By locating harm directly in the norms created, McGowan effectively offers another mechanism by which hate speech functions in the two ways Lawrence thinks it does. When a speaker enacts in a conversational setting the norm that members of targeted groups ought to be discriminated against, this diminishes targets' status. It is now a local norm that targets do not get equal treatment, that their interests do not receive equal weight, and therefore that they do not possess equal social and political status. The integration of this norm into the conversational score is a manifestation of targets' diminished dignity. Moreover, these norms exclude targets from public accommodations. Norms calling for exclusion do not merely galvanize others to exclude, but themselves communicate to targets that they are unwelcome. It would be odd to say, analogously, that a restaurant posting a 'Whites Only' sign merely *causes* discrimination or exclusion, as if these were disparate effects at the end of a causal chain. Rather, the restaurant simply discriminates and excludes, harming non-white members of society. The norms themselves make members of targeted groups unwelcome.

The way that racist hate speech enacts discriminatory norms gives us further reason to reject the contention that hate speech merely expresses the opinion of one individual and has minimal effect on the dignity of marginalized groups. These norms constitute a kind of non-material 'Whites Only' sign that exists in the conversational score rather than on the wall. If a white person on a public bus hung a literal 'Whites Only' sign, this would not only express their opinion, but deny non-white folks equal status and access to public accommodations. Similarly, the discriminatory norms enacted by hate speech not only tell targets that one person deems them to be second-class citizens, but serve as a mechanism that denies them their dignity and access and makes them second-class citizens. Members of marginalized groups know they are unwelcome and that they will be subject to abuse and perhaps even violence if they violate the norms.

39. McGowan, *supra* note 35 at 23.
40. *Ibid* at 176.

https://doi.org/10.1017/cjlj.2021.18 Published online by Cambridge University Press

The other way that hate speech leads to harm for McGowan is causal: by marking racial discrimination as acceptable or obligatory, the norms hate speech enacts cause others to discriminate against its targets. In this way, hate speech in effect *causes harm by constituting harm*, compounding its illocutionary discrimination with damaging perlocutionary effects. This function of hate speech again reflects the way that hate speech harms beyond merely expressing one person's opinion, motivating uptake of discriminatory behaviors by others in society. Moreover, this causal account also avoids the objection from subjective offense, as it focuses only on speech that produces a particular type of harm in a particular way, i.e. speech that causes discrimination by enacting discriminatory norms.

Some might think that the causal part of this theory offers an implausible account of the responses hate speech actually tends to elicit. Like Simpson, one might argue that hate speech is much more likely to inspire derision and criticism than uptake and support. Yet this seems to overlook the group dynamics often involved in racial insults. One need look no further than the 'counter-protests' that have recently sprung up all over the U.S. in response to Black Lives Matter protests. These counter-protests tend to start with older white people heckling protestors with phrases like "All Lives Matter" or "Blue Lives Matter," but as they've gone on, many protests have devolved into expressions of racial insults and violence. In one instance, a group of older white men started screaming "White Power" at passing protestors,[41] and in another, a group of men surrounded and then punched a young woman carrying a "Black Lives Matter" sign.[42] It seems perfectly plausible that the phenomenon McGowan describes has been at work in these and other instances: counter-protestors are emboldened by the speech of those around them, gaining a sense that racial insults and violence are acceptable in this context. These hecklers feed on each other, becoming comfortable with increasingly vile racial language, until uttering hate speech, treating marginalized groups as inferior and unworthy of respect, and even performing violence becomes socially allowable, if not expected.

The U.S has also seen the phenomenon McGowan highlights at play on a much broader, national scale. Over the last several years, there has been much discussion over the racist rhetoric employed by former President Donald Trump and concern that such rhetoric has led to more racist discourse among the masses. Recent scholarship indicates that there is truth to this suspicion.

A 2018 study from political scientist Brian Schaffner shows evidence for what

41. See Benjamin Swasey, "Trump Retweets Video Of Apparent Supporter Saying 'White Power'" (28 June 2020) NPR, online: https://www.npr.org/sections/live-updates-protests-for-racial-justice/2020/06/28/884392576/trump-retweets-video-of-apparent-supporter-saying-white-power.

42. See Kassidy Vavra, "'RIPPED MY SIGN AWAY' BLM protester, 19, 'thrown against wall and punched in head' as she's surrounded by counter-demonstrators" (18 June 2020) The US Sun, online: https://www.the-sun.com/news/1002296/blm-protester-thrown-punched-head-surrounded-counter-demonstrators.

https://doi.org/10.1017/cjlj.2021.18 Published online by Cambridge University Press

he calls the "Trump Effect": exposure to President Trump's prejudiced statements during the 2016 campaign made people more likely to write offensive things both about groups Trump has targeted and other identity groups.[43] A 2020 study from a group of political scientists reveals a similar trend, showing that in the absence of prejudiced speech from elites, prejudiced citizens tend to keep quiet, but when elites express prejudice, prejudiced citizens are more likely to follow suit.[44] Both of these studies highlight the causal effects of hate speech that McGowan discusses, i.e. that it makes expressions of racial animus seem socially acceptable or even obligatory, motivating others to utter racial insults and treat members of marginalized groups with less respect.

One might object that this process does not actually perpetuate racial hierarchies because it merely leads those who are already prejudiced to express that prejudice, rather than driving prejudiced attitudes. Even if this were true, the constitutive element of McGowan's account shows that an increase in racist speech can itself further malign the dignity of targeted groups, even without an increase in racist attitudes. By diminishing their status and excluding marginalized racial groups from the privileges of full citizenship, hate speech itself contributes to racial hierarchy. Yet there's reason to think that the norms enacted by racist hate speech can also motivate racist attitudes, because an environment in which such speech abounds presents racism as a socially acceptable option. As Waldron puts it, advocates of hate speech regulation fear that as a result of persistent racial insults, "ordinary people will think and act on the assumption that the place of minority members in ordinary life is up for grabs."[45] We can imagine young people in such an environment, still in the process of developing their views about the world, coming to see racism not as deplorable, but as a live option. Social pressure certainly has an effect on the views we embrace, and if people see that others in their community express racist attitudes, they'll be more likely to think such attitudes are acceptable and give them serious consideration. Racist expression may also harden the views of those who already have racist proclivities. Living in a community in which racist expression is commonplace will lead one to be more complacent in one's racist views, while consistent criticism of such views will force one to grapple with and question them.

Attention to the ways that racist hate speech enacts discriminatory norms that others follow therefore gives us further reason to accept Waldron's argument that it denigrates targets' status and ought to be regulated. By enacting such norms, hate speech excludes members of targeted groups and encourages others to treat them with dehumanizing animus and disregard, producing status harms that

43. See Brian Schaffner, "Follow the Racist? The Consequences of Trump's Expressions of Prejudice for Mass Rhetoric" (2018) Semantic Scholar, online: https://www.semanticscholar.org/paper/Follow-the-Racist-The-Consequences-of-Trump-%E2%80%99-s-of-Schaffner/3ff29822155973661029da17a4c0610088e15340#paper-header.

44. See Benjamin Newman et al, **"**The Trump Effect: An Experimental Investigation of the Emboldening Effect of Racially Inflammatory Elite Communication" (2021) 51:3 British Journal of Political Science 1138.

45. Waldron, *supra* note 11 at 153.

https://doi.org/10.1017/cjlj.2021.18 Published online by Cambridge University Press

outweigh the relevant free speech interests. Again, I do not doubt that such speech implicates speakers' expressive, deliberative, or political interests and that regulation is therefore subject to a higher standard of evidence, but contend that the harm it does outweighs those interests and meets that standard. For, as I outlined in Section 4, this harm violates the guarantee that all members of society enter social and political life as equals, arguably the fundamental requirement of liberal politics, and threatens targets' other basic rights.

7. Why Not the CLS Approach?

One might wonder why, if we're employing resources from the CLS approach to shore up the dignity approach, we don't simply employ the CLS approach. The reason is that the CLS approach has a much different aim than the dignity approach. Rather than attempting to show that the harms of racist hate speech outweigh its benefits, proponents of the CLS approach maintain that racist hate speech is unrelated to the core interests underlying free expression and is therefore uncovered. In this section, I will argue that this account ultimately fails, showing that hate speech ultimately does promote at least two of the most prominent interests commonly cited in favor of the FSP.

a. Hate Speech and the Deliberative Interest

Proponents of the CLS approach contend that hate speech does not promote what Joshua Cohen calls the deliberative interest: the interest people have in accessing valuable information that can help them lead better lives and help others do the same. Lawrence argues, "The racial invective is experienced as a blow, not a proffered idea" and concludes that hate speech ought to be considered the "functional equivalent of fighting words."[46] Delgado echoes this sentiment when he says that racial insults "are not intended to inform or convince the listener."[47] The idea is that racist hate speech has the intent and impact of injuring its victims and promoting fear, intolerance, and violence, rather than sharing an idea or motivating useful discourse. In its strong form, this argument insists that hate speech lacks what I'll call intellectual content, i.e. an articulated idea or argument, but is rather akin to a "blow," as Lawrence describes it. Yelling a racial slur at someone does not express an idea and is unlikely to spark a conversation, let alone a thoughtful and productive one. On the contrary, such speech is more likely to silence and drive away its targets.

Yet the position that racist hate speech lacks intellectual content is inconsistent with critical legal theorists' identification of the harm in hate speech, because these thinkers insist that it is the message conveyed by hate speech that makes it so pernicious. As one example of hate speech, Lawrence, Matsuda, Delgado,

46. Lawrence, *supra* note 23 at 68.
47. Richard Delgado, "Words That Wound: A Tort Action for Racial Insults, Epithets, and Name Calling" in Matsuda, *supra* note 23 at 108.

https://doi.org/10.1017/cjlj.2021.18 Published online by Cambridge University Press

and Crenshaw discuss a dispute among white students and a Black student at Stanford University over Beethoven's heritage, after which a group of white students posted in image of Beethoven with Sambo-like features on the Black student's door. Regarding this incident, the authors explain, "The message said, 'This is you. This is you and all of your African-American brothers and sisters. You are all Sambos. It's a joke to think you could ever be a Beethoven. It's ridiculous to believe that you could ever be anything other than a caricature of real genius.'"[48] It is clear in this case that the authors think it is the idea expressed, the idea that Black people are inferior and incapable of achieving the feats of genius achieved by white people, that makes this speech so harmful. The speech here is not merely experienced as "a blow" as Lawrence suggests, but as an idea.

This example is not an aberration in the CLS account, but rather reflects its central idea. Delgado stresses that it is the internalization of the message of inferiority expressed by hate speech that is so damaging to its victims, as many "come to believe the frequent accusations that they are lazy, ignorant, dirty, and superstitious."[49] He cites psychologist Kenneth Clark, who explains that, "Human beings ... whose daily experience tells them that almost nowhere in society are they respected and granted the ordinary dignity and courtesy accorded to others will, as a matter of course, begin to doubt their own worth."[50] It is the idea it expresses that makes hate speech so vicious, the idea that members of marginalized racial groups are inferior and unworthy of equal status. As Henry Louis Gates notes, this means that racist hate speech must have intellectual and political content. "[I]f Delgado and his fellow contributors have a central message to impart," Gates explains, "it's that racial insults are profoundly political."[51] It is the disturbing political message expressed by hate speech that makes it more emotionally and psychologically affecting than a scream or incitement to violence. If it truly lacked intellectual content, hate speech would lack the poignancy that makes it such a compelling candidate for regulation.

Of course, there may be some expressions that incite fear and make targets feel unwelcome that do not express a political message. Screaming inarticulately at someone on the subway or blocking the path of people of color as they walk into a restaurant may stoke fear or incite violence without expressing a clear political idea. Such expressions exclude not by expressing a message of inferiority, but rather by threatening their targets directly. Yet Matsuda, Delgado, and Lawrence are not exclusively concerned with this narrow range of expressions. Rather, they focus particularly on those utterances that do express a message of inferiority, that exclude by telling targets they are unworthy of equal recognition. The prior category of expressions—those that threaten rather than express ideas—would already enjoy regulation under prohibitions on incitement to

48. Charles Lawrence III et al, "Introduction" in Matsuda, *supra* note 23 at 8.
49. Delgado, *supra* note 47 at 91.
50. *Ibid*.
51. Henry Louis Gates Jr, "War of Words: Critical Race Theory and the First Amendment" in Gates Jr et al *Speaking of Race, Speaking of Sex: Hate Speech, Civil Rights, and Civil Liberties* (NYU Press, 1995) at 26.

https://doi.org/10.1017/cjlj.2021.18 Published online by Cambridge University Press

violence or direct threats. The racial nature of such utterances seems almost superfluous, for it is the threatening and dangerous environment they create that is crucial, not the message expressed. Indeed, if such expressions are to lack intellectual content, their racial content *has to be* superfluous. Critical race theorists like Matsuda, Delgado, and Lawrence are rightly concerned with something entirely different, a category of expressions that perpetuates racial hierarchy and prejudice by promoting a message of inferiority. Here the messages' intellectual content is exactly what makes them so harmful. Indeed, Gates argues that the harmfulness of hate speech is often inversely related to its vulgarity and thoughtlessness. The sterile assertion that one is inferior and unworthy of equal status in society is often much more devastating than the angry use of a racial slur.[52]

One might try to sidestep these objections by presenting a weaker version of the view that hate speech has no essential connection to the acquisition of truth. This version says that it's not that such speech does not express an idea, but rather that it's not intended to—and is unlikely to—spark productive conversation. Lawrence at one point seems to endorse this weaker view when he argues that in the wake of hate speech, "it is unlikely that dialogue will follow. Racial insults are undeserving of first amendment protection because the perpetrator's intention is not to discover truth or initiate dialogue, but to injure the victim."[53] It is not that such speech lacks any intellectual content, but rather that neither the intent nor impact of the speech is to spark deliberation that's likely to uncover the truth.

And yet, that one thinks such speech unlikely to lead to the acquisition of truth doesn't seem sufficient for it being uncovered by a free speech principle. If I utter to a man on the street, "The frogs on the moon are triangles, not circles," this seems much more likely to inspire befuddlement than productive conversation, and yet we would certainly not say such speech should be uncovered by a free speech principle. Of course, such an utterance is also unlikely to do as much harm as hate speech, but the question of harm is not relevant to whether or not an utterance is covered, but only to whether if, uncovered, it ought to be regulated. The reason we would extend coverage to my utterance on the street even though it seems to lack deliberative value is that it is not our place to judge in advance which ideas are likely to promote the acquisition of truth. This is at the very core of the free speech justification: we must tolerate all ideas because we don't know in advance which are true, and even false or absurd ideas may contribute to knowledge by provoking response. By silencing even false opinions, society loses, according to Mill, "the clearer perception and livelier impression of truth, produced by its collision with error."[54] This does not mean that no speech can be regulated, but rather that all speech which expresses an idea is relevant to the deliberative interest.

The efforts of philosophers of language to characterize the linguistic function of hate speech fare no better in showing it lacks intellectual content. In an earlier

52. *Ibid* at 46-47.
53. Lawrence, *supra* note 23 at 68.
54. John Stuart Mill, *On Liberty* (WW Norton & Co, 1975) at 18.

https://doi.org/10.1017/cjlj.2021.18 Published online by Cambridge University Press

work, Maitra and McGowan argue that, like other types of speech uncovered by a plausible free speech principle, expressions of hate speech are "significantly obligation-enacting utterances."[55] This means that such speech creates obligations of considerable consequence, in that they are not easily met or there are serious repercussions for not meeting them. Maitra and McGowan argue that such utterances are uncovered because, while the primary value of speech is in its deliberative function, the principal effect of significantly obligation-enacting utterances is to make changes in the world, i.e. creating obligations. Other examples include speech which initiates or dissolves contracts, conspires to commit a crime, or falsely advertises—none of which receive the heightened protections reserved for expression.[56] Racist hate speech is significantly obligation-enacting by Maitra and McGowan's lights because it enacts obligations to do something illegal, i.e. to deny targeted groups equal access to public spaces, and because there are social costs for refusing to fulfill the obligation.[57]

The problem with Maitra and McGowan's assertion that utterances of racist hate speech ought to be uncovered because they are "significantly obligation-enacting utterances" is that many such utterances have intellectual and political content. Take for instance Martin Luther King's famous assertion in his Letter from a Birmingham Jail that "[o]ne has a moral responsibility to disobey unjust laws."[58] This is clearly a significantly obligation-enacting utterance as it asserts an obligation to break the law, which Maitra and McGowan explicitly identify as a significant obligation. And yet, it also clearly has deliberative value, helping people work out how they ought to act when faced with institutionalized injustice.[59] It is therefore clear that not all significantly obligation-enacting utterances lack expressive or deliberative value, and therefore that not all such utterances should be uncovered by the FSP.

b. Hate Speech and the Expressive Interest

Critical legal theorists also reject the connection of hate speech to what Joshua Cohen calls the 'expressive interest': a "direct interest in articulating thoughts, attitudes, and feelings on matters of personal or broader human concern."[60] Cohen's invocation of the expressive interest is founded on the intuition that part of living a good life includes the ability to bear witness to what one

55. Ishani Maitra & Mary Kate McGowan, "On Racist Hate Speech and the Scope of a Free Speech Principle" (2010) 23:2 Can JL & Jur 343 at 353.
56. *Ibid* at 345.
57. *Ibid* at 369.
58. Martin Luther King Jr, "Letter from the Birmingham Jail" (1963) 212:2 *The Atlantic Monthly* 78.
59. One might object that this assertion doesn't enact a new obligation, but merely affirms an existing obligation by communicating a moral truth. Yet by making this moral truth known to listeners, this utterance does present them with a new obligation, just as I might enact a new obligation by telling someone that a clothing company employs child labor. If previously ignorant of this fact, the listener would not have had an obligation to avoid patronizing this company before hearing me, but would have the obligation afterwards.
60. Cohen, *supra* note 6 at 224.

https://doi.org/10.1017/cjlj.2021.18 Published online by Cambridge University Press

perceives as true and valuable. As Delgado puts it, individuals have a right "to develop their full potentials as members of the human community."[61] Yet Delgado also maintains that expressing hate speech plays no central role in promoting this interest. Rather, he argues that bigotry instead stifles one's moral and social growth, precluding one from achieving higher values like justice or inclusion. Moreover, he does not see racial insults primarily as vehicles for self-expression, but rather as attempts to injure or insult. Just as the intent and impact of hate speech is not to promote discussion, neither is it to express oneself.[62]

This reasoning is flawed on two fronts. For one, the expressive interest does not only extend to expressions likely to promote one's intellectual or moral development. Rather, as Cohen argues, the relevance of the expressive interest depends on an individual's subjective stance towards the topic they wish to address.[63] What is important is that one takes oneself to have a strong obligation to express some view or sees the issue as important to justice or human welfare. The expressive interest concerns individuals' ability to bear witness. There may be plenty of views that individuals see as important to express that we from the outside might view as objectively unimportant or even deplorable and dangerous. This seems to be the case with hate speech: although we may judge the ideas expressed to be objectively deplorable, racists see them as crucially important to share. As Jeremy Waldron argues, the ideas that purveyors of racist hate speech express "are the very messages, out of all things a person could express, which matter most to them."[64] Because of their stance towards such issues, the expressive interest is implicated in hate speech, even if such expression is unlikely to promote speakers' intellectual or moral development and instead likely to stunt it.

The other problem with Delgado's argument is the impropriety of making government the arbiter of what views promote intellectual or moral development. It would be odd, for instance, if protections did not extend to reality TV or romantic comedies because the government deemed them of little intellectual value, but did extend to documentaries or historical dramas. Monitoring the quality of our expression and extending protections only to that expression deemed to possess intellectual or moral value simply does not seem like a proper role for government.

Protecting the dignity of its citizens from racist abuse is, on the other hand, a proper role of government. Of course, allowing a government to determine what expressions denigrate targets is not without its dangers. It is certainly possible that a government would abuse this power in order to limit expression it doesn't like. Yet the justification of government actions here matters: while the government does not have the authority to determine what expression has intellectual value, it does have the authority to preserve dignity. Risks of abuse are tolerable when they are a side effect of a government doing its legitimate duty to protect the equal status of citizens, but not when they are a product of an already

61. Delgado, *supra* note 47 at 108.
62. *Ibid* at 108.
63. Cohen, *supra* note 6 at 225.
64. Waldron, *supra* note 11 at 149.

https://doi.org/10.1017/cjlj.2021.18 Published online by Cambridge University Press

questionable practice. Moreover, allowing a legislature to limit expression that threatens dignity grants the government a more circumscribed power and creates fewer opportunities for abuse. While a government might deem any expression it doesn't like to have no intellectual value, there are fewer candidates for speech that erodes dignity. For instance, such regulations could only extend to direct attacks on people, not challenges to policies or ideas.

8. The Scope and Nature of Regulation

My aim in this paper has been to establish that at least certain types of hate speech ought to be regulated because such speech erodes targets' dignity, not to outline the precise scope or form of regulation. I take it that some kinds of speech quite clearly threaten targets' dignity and ought to be regulated, especially extreme expressions of racial discrimination that mark members of marginalized groups as severely morally or rationally deficient. If a white man calls a Black woman a slur on a public bus and declares to onlookers that she and other Black people ought to be rounded up and killed, he pretty obviously denies her and other Black people recognition respect, precludes her from enjoying the full privileges of citizenship, and pollutes the social environment with discriminatory norms.

Yet other cases will surely provoke greater controversy, which warrants some brief remarks on the specific nature of regulation. One question concerns the arenas in which governments ought to regulate speech: should regulations extend only to spoken hate speech, or also to speech posted online, or published in a book or newspaper? Another is where to draw the line between politically acceptable arguments and extreme expressions of racial discrimination that warrant regulation. For instance, Gates gives an example of a professor telling a Black student that it's not the student's fault if he struggles in his classes, because he's only been admitted thanks to an affirmative action policy that places underprepared students into demanding educational environments.[65] Would this count as racist hate speech for the purpose of regulation?

I am willing to bite the bullet with respect to what some might deem undesirable implications. Online speech and published materials certainly can denigrate the status of marginalized groups, and therefore I think it's perfectly plausible to regulate white supremacist internet trolls or books arguing that members of marginalized groups ought to be eliminated. The most obvious place-based exceptions would cover private conversations or writings, never distributed to a wider audience, which either will not have the effect of eroding targets' dignity or will introduce privacy considerations that outweigh their minimal effects on dignity. As for what specifically counts as an extreme enough expression of racial discrimination to warrant regulation, this is a complex question I cannot address here. What I will say is that there are certainly racially controversial types of speech that do not call for regulation, including arguments regarding racially-loaded political topics like police reform or affirmative action. Whether speech

65. Gates Jr, *supra* note 51 at 46.

https://doi.org/10.1017/cjlj.2021.18 Published online by Cambridge University Press

on one of these topics creates dignitary harms outweighing its value as speech depends on its specific content, delivery, and audience. For instance, an op-ed arguing that affirmative action tends to hurt students who grow up with fewer resources and support would not count. Telling a Black student that affirmative action unjustifiably privileges inherently inferior Black applicants would.

The nature of such regulations would likely be similar to those instituted in the U.S. by the *Civil Rights Act*.[66] Civil rights violations, for instance refusing to serve someone at a restaurant because of their race or religion, carry punishments ranging from fines to imprisonment. Given that the longer sentences are reserved for violations involving bodily injury,[67] punishments for hate speech would likely land on the lower end of this spectrum, comprising mostly fines and perhaps short prison sentences in particularly egregious cases.

9. Conclusion

Taking stock, this paper fills in the gaps in Waldron's dignity approach to hate speech, showing how hate speech does in fact erode the dignity of its targets by promoting racial hierarchies. I borrow insights from the critical legal studies (CLS) approach into the ways that hate speech erodes its targets' dignity: first, by constituting discrimination in virtue of denying them equal status and the full privileges of citizenship, and second, by enacting discriminatory norms that promote racial animus against targeted groups. On my view, this proclivity of racist hate speech to erode its targets' dignity outweighs the free speech interests that hate speech promotes, making such speech unprotected by the FSP and justifying regulation. However, I ultimately reject the CLS justification of hate speech regulation because it misinterprets the relation between hate speech and the interests underlying free expression and underestimates the scope of these interests. I therefore conclude that while hate speech does indeed promote the deliberative and expressive interests underlying free expression, these interests are outweighed by hate speech's proclivity to strip its targets of their dignity. As a result, hate speech is covered by the FSP, but ultimately unprotected, and therefore ought to be regulated.

Acknowledgments: I would like to thank Aatif Abbas, Ugur Altundal, Luvell Anderson, Kenneth Baynes, Mary Kate McGowan, Keshav Singh, Robert Van Gulick, Hannah Widmaier an anonymous referee at the CJLJ, and audiences at Syracuse University, UCLA, and USC for their helpful comments on previous versions of this paper.

Chris Bousquet is a PhD student in philosophy at Syracuse University. He works primarily in legal and political philosophy, especially on the topics of free expression and its limits, democratic theory, and the ethics of work. Email: cbousquet422@gmail.com.

66. *Civil Rights Act of 1964*, 42 USC § 1971 et seq (1988).
67. "Federal Civil Rights Statutes" (16 May 2016) online: https://www.fbi.gov/investigate/civil-rights/federal-civil-rights-statutes.

https://doi.org/10.1017/cjlj.2021.18 Published online by Cambridge University Press