Noxious
Annu. Rev. Political Sci. 2019. 22:X–X
https://doi.org/10.1146/annurev-polisci-051517-012343

Free Speech and Hate Speech

Jeffrey W. Howard

Department of Political Science, University College London, London WC1H 9QU, United Kingdom; email: jeffrey.howard@ucl.ac.uk

Keywords

free speech, hate speech, autonomy, democracy, defamation

Abstract

Should hate speech be banned? This article contends that the debate on this question must be disaggregated into discrete analytical stages, lest its participants continue to talk past one another. The first concerns the scope of the moral right to freedom of expression, and whether hate speech falls within the right's protective ambit. If it does, hate speech bans are necessarily unjust. If not, we turn to the second stage, which assesses whether speakers have moral duties to refrain from hate speech. The article canvasses several possible duties from which such a duty could be derived, including duties not to threaten, harass, offend, defame, or incite. If there is a duty to refrain from hate speech, it is yet a further question whether the duty should actually be enforced. This third stage depends on pragmatic concerns involving epistemic fallibility, the abuse of state power, and the benefits of counter-speech over coercion.

INTRODUCTION

Does a proper commitment to freedom of expression demand the legal protection of so-called hate speech? The world's democracies fiercely disagree on the answer to this question. Consider the United Kingdom, where it is a criminal offense to incite racial or religious hatred (Brown 2016). While details differ, legislation of this sort exists in the preponderance of developed

democracies, including Australia, Denmark, France, Germany, India, South Africa, Sweden, and New Zealand (Waldron 2012, p. 8). Yet such legislation would clearly be struck down as unconstitutional in the United States, as an affront to free speech. The US Supreme Court has held uncompromisingly that laws perpetrating discrimination on the basis of viewpoint, especially those that suppress the expression of certain moral and political convictions, violate the First Amendment (Stone 1987, 1994). That is so even when the rationale for the suppression is to prevent any criminal violence that hateful speech might inspire [*Brandenburg v. Ohio* 395 US 444 (1969)].

It is tempting to view this standoff as a matter of whether freedom of speech should be granted priority over other political values. "Although free speech is an important value," writes Parekh (2012, p. 45), "it is not the only one." This common suggestion is that our commitment to free speech must be balanced when its demands conflict with other normative commitments, such as the social equality, dignity, or security of historically marginalized citizens. This way of framing the debate is resonant, both politically and philosophically (Fish 1994; Waldron 2012, pp. 145–147, 172). Yet it is a defective frame, for two reasons. First, if this were truly a case of ineliminable moral conflict that could be resolved only through balancing, the question of what value should win out in each case seems either indeterminate or dependent on the assertion of our divergent intuitions as to the relevant values' differential weights. A satisfying treatment of this dispute should aspire to do more than recommend ad hoc balancing; it should at least attempt to identify a coherent and well-supported theoretical position that can offer clear normative guidance.

Second, and more importantly, the balancing model mischaracterizes the real debate. It suggests that those who oppose bans on hate speech are the real defenders of free expression, whereas those who support bans are hostile to free speech, or at the very least comfortable with infringing it for the sake of other values. Yet the crucial debate is not about whether we should infringe free speech in order to stamp out hateful attitudes and the various evils they engender. Rather, it is about whether hate speech even constitutes the sort of expression that the right to freedom of expression exists to protect, or whether it instead falls outside that right's ambit of protection.

That suggestion recommends a promising methodology for how to reason our way through this dispute: identify the arguments that purport to justify the moral right to freedom of

expression, and then assess whether these arguments count in favor of including hate speech within the protection of that right. Applying this methodology, I rehearse the most influential scholarly arguments that defend the legal protection of hate speech. I argue that if hate speech is protected by the moral right to freedom of expression—something I take no firm stand on here—then it follows that bans on hate speech are unjust. However, if hate speech is not protected, it is a further question whether bans on it are justified. Demonstrating that conclusion depends, in part, on showing that there is a moral duty to refrain from hate speech, which laws banning hate speech serve to enforce. I canvass potential candidates for this duty in the scholarly literature on hate speech, focusing on putative duties not to threaten, harass, offend, defame, or incite. Finally, I argue that even if there is a duty to refrain from hate speech, it is yet a further question whether that duty should actually be enforced. I consider various reasons why it might be misguided to enforce this duty, including concerns over epistemic fallibility, the abuse of state power, and the preferability of counter-speech over coercion. While these considerations are typically proffered by scholars as justifications for the right to free speech itself, they are best conceived, I argue, as reasons why the duty to refrain from hate speech should not be legally enforced.

While I will intimate what I take to be the strengths and weaknesses of particular arguments, my chief goal in what follows is not to convince readers of any particular position. Instead what I offer is an analytical framework that organizes the current scholarly debate, so that readers are equipped with an intellectual structure they can use to make up their own minds. The nerve of my contribution is that the largely muddled debate over hate speech needs to be broken down into discrete analytical stages—clearly distinguishing (*a*) the protective scope of the moral right to free speech and whether hate speech falls within it; (*b*) the demands of any duties we have to refrain from hate speech; and (*c*) various further arguments concerning the risks of actually enforcing these duties, which become relevant when designing constitutional or statutory free speech protections.

DEFINITIONS

Before we proceed, definitional comments on free speech and hate speech are in order.

Free Speech

Any invocation of the term free speech may be referring to any one of a variety of meanings. One is the moral right to freedom of expression, a fundamental moral requirement that agents be free to express themselves and communicate with others.[1] This right generates correlative negative duties on the part of other agents. It also generates positive duties on the part of the state to protect those negative duties, e.g., by providing police protection for threatened speakers (Scanlon 2011, p. 332).

A second meaning is the legal right to freedom of expression that is entrenched in the law of a given jurisdiction, as with the First Amendment to the US Constitution. Normative political theorists tend to assume that this second is, or at least ought to be interpreted as, simply the legal codification of the first. But as we will see, concerns about human fallibility and institutional design may mean the exact contours of the moral right to free speech should not be precisely mirrored in the law (see Shiffrin 2003, p. 1184). In what follows, I discuss both the moral and legal rights to free speech.

Hate Speech

There is enormous variation in the definition of hate speech. Clearly it does not merely refer to speech that expresses hatred; after all, it is fully appropriate, we might think, to express hatred at heinous injustice, for example (Post 2009, p. 123). Instead hate speech is a term of art, referring to the particular expressions of hatred against particular (groups of) people in particular contexts. For example, Germany's laws restrict expression that affronts "the human dignity of others by insulting, maliciously maligning, or defaming segments of the population," and New Zealand bans speech with "threatening, abusive, or insulting…words likely to excite hostility against or bring into contempt any group of persons…on the ground of the color, race, or ethnic or national…origins of that group of persons" (Waldron 2012, p. 8). Debates rage on how, exactly,

---

[1] For present purposes, I set aside skeptical views toward the existence of a special moral right to freedom of expression (as distinct from a general liberty principle), as expressed by, e.g., Schauer (1982, 1983, 2015) and Alexander (2005). I am convinced by Kendrick's (2017) recent defense.

to define hate speech (Brison 1998, p. 313; Brown 2017a,b; Matsuda 1989, pp. 2357–58; Quong 2011, p. 305n; Waldron 2012, pp. 8–9), as well as on what sorts of vulnerable groups are its paradigmatic targets (Brown 2017c,d).

In what follows, I adopt the three-pronged characterization of hate speech offered by Parekh (2012, p. 41), since it subsumes the concerns of numerous authors. On his view, hate speech, first, "is directed against a specified or easily identifiable individual or, more commonly, a group of individuals based on an arbitrary or normatively irrelevant feature"; second, it "stigmatizes the target group by implicitly or explicitly ascribing to it qualities widely regarded as undesirable"; and third, it casts the "target group…as an undesirable presence and a legitimate object of hostility".

FREE TO HATE

Why think that the moral right to freedom of expression protects hate speech? Following Scanlon (1979), one promising methodology for exploring questions like this is to identify the considerations that serve to justify the moral right to freedom of expression in the first place, and then inquire whether a controversial category of speech is sufficiently related to these considerations to qualify for protection (see also Greenawalt 1989, Yong 2011). I apply that methodology here by reviewing the considerations widely believed to underwrite the moral right to free speech, which in turn justify a corresponding legal right.

I do not address every argument for free speech. For example, two decades ago it would have been unthinkable to ignore the idea of an unregulated marketplace of ideas, and its importance for the discovery of truth, as a central argument for free speech, an argument with influential roots in the work of J.S. Mill [1978 (1859)]. Yet the idea that Mill even endorsed a view of this kind, according to which the truth is simply fated to prevail through the open clash of ideas, is now highly controversial (principally due to Gordon 1997). Exegetical concerns aside, the general empirical claim on which this model appears to rest has been thoroughly discredited.[2] So,

[2] For just a sample of the criticisms the view has received, see Alexander (2005, p. 128), Bambauer (2006), Brietzke (1997), Glaeser & Sunstein (2014), Ingber (1984), Schauer (2010), Simpson (2013, p. 722), Sparrow & Goodin (2007), and Waldron (2012, p. 156). As always there

while we need not deny that the idea of a marketplace of ideas may play some role in the justification of free speech, it not the central argument for it. For the sake of space, I restrict myself to four other arguments for free speech: the argument from listener autonomy; the argument from speaker autonomy; the argument from democracy; and a new thinker-based argument.

The Argument from Listener Autonomy

I consider first the view that the moral right to freedom of expression (hereafter "free speech") is required by respect for the autonomy of listeners. "To regard himself as autonomous," writes Scanlon (1972, pp. 215, 217), "a person must see himself as sovereign in deciding what to believe and in weighing competing reasons for action." As a result, he concludes: "The harm of coming to have false beliefs is not one that an autonomous man could allow the state to protect him against through restrictions on expression." Dworkin (1996, p. 200) agrees: "Government insults its citizens, and denies their moral responsibility, when it decrees that they cannot be trusted to hear opinions that might persuade them to dangerous or offensive convictions."

This argument seems to provide a powerful case for opposing bans on hate speech (see also Nagel 2002, p. 44; Stone 1987, pp. 56–57). Yet the criticisms of this view are significant. Set aside the possibility that this view is not, in fact, a successful argument for free speech as a general matter, as Scanlon (1979, pp. 532ff) came to believe. Even accepting that free speech is at least partly justified by concerns of listener autonomy, the claim that such concerns should move us to protect hate speech is weak.

Three arguments here are worth pointing out. First, while some hate speech takes the form of inciting listeners to adopt hateful views, and so arouses the concern of this argument, plenty of hate speech—e.g., "face-to-face vilification" in which hateful speakers directly harass, bully, intimidate, and threaten vulnerable citizens—does not take this form (Brison 1998, p. 328).

Second, even if citizens *qua* listeners are concerned that the state respect their autonomy, they are also surely concerned *qua* potential victims that the state protect them from the kinds of discrimination and violent attack that hate speech is feared to inspire. If we have a social world

---

are exceptions; see, e.g., Marshall (1995) and Volokh (2011).

in which hate speech is widely protected, citizens are accordingly rendered more vulnerable to the wrongdoing that such speech inspires. It is wholly coherent to suppose that autonomous, democratic citizens could instruct the government to protect them from such wrongdoing by outlawing speech that inspires it (Amdur 1980, p. 299; Brison 1998, p. 329; Greenawalt 1989, p. 115).

Third, the argument from listener autonomy applies paradigmatically to expression with propositional content—i.e., expression that offers reasons for what listeners ought to believe or do, which (the argument goes) they are entitled to evaluate for themselves. But not all expression advances propositional content to be evaluated cognitively. Among the most significant arguments against pornography, especially violent pornography, is that it operates subrationally in that it potentially "alters its users' desires or beliefs by nonpersuasive means" (Scoccia 1996, p. 784). Through a process of subliminal suggestion, violent pornography might induce its consumers to acquire unconsciously a certain mental state, even if the consumers would, on reflection, be horrified to possess such a state (Scoccia 1996, p. 787; see also Sunstein 1986, pp. 607–8). Indeed, many have argued that pornography, at least in an aggressively misogynistic form, itself constitutes hate speech (see especially MacKinnon 1996). If so, and if the mechanism by which such hate speech influences is one that bypasses, rather than engages, agents' autonomous capacities, then its suppression is conceivably consistent with respect for listener autonomy.

The Argument from Speaker Autonomy

A second argument for free speech is grounded in the autonomous interests of speakers. The autonomy interest on which I will focus is the interest that citizens have in framing, revising, and pursuing a conception of a good life—what Rawls (2005, pp. 47ff) calls "the second moral power." Free speech is indispensable to the development and exercise of this vital power. Agents who have this power have an expressive interest in testifying to others about what they take to be the best way to live, and in endeavoring to persuade others to adopt their views (Cohen 1993, p. 224). They also have a deliberative interest in enjoying circumstances that are propitious for deliberation about the good life, which depends on free communication with others in conditions

of full information (Cohen 1993, pp. 228–29).[3] Without free speech, the capacity to be a robust author of one's own life—for people to pursue their own good in their own way, to paraphrase Mill—is seriously stunted. As Baker (1978, p. 991) puts it, a central purpose of free speech protections is to "delineate a realm of liberty for self-determined processes of self-realization."

If free speech is justified by our autonomy interest in pursuing a good life, does it thereby protect hate speech? One argument that it does *not* holds that even if autonomy is generally valuable, it lacks value when deployed in the service of seriously mistaken ends (Raz 1991, p. 319; cf. Waldron 2012, pp. 146–47). But this suggestion may be too quick. Suppose we claimed to respect someone but only let him speak when the views he was expressing were correct. It seems that respecting people requires letting them express themselves even when they are wrong. To deny this would conflate what Darwall (1977) famously called recognition respect with appraisal respect. "Respect for personhood, for agency, or for autonomy," writes Baker (1997, p. 992), "requires that each person must be permitted to be herself and to present herself." For this requirement to be robust, it cannot evaporate the moment a speaker's autonomous expression becomes distasteful, or even when it risks harm. A citizen with hateful views must therefore be permitted to share those views, simply because they constitute the citizen's "view of the world" and so the speech "expresses her values" (Baker 2009, p. 143; cf. Schauer 2015, p. 128).

One reply to this line of argument is to adopt the balancing model I criticized above—to concede that there is indeed value in expression of this sort but to deem it outweighed by countervailing harms (Waldron 2012, pp. 165–66, 172). After all, as is widely recognized, people express who they are through their objectionable actions (e.g., killing those they hate), as well as their objectionable words. So perhaps this expressive value is always present in speech and nonspeech cases alike and always simply needs to be weighed against the costs.

Another line of reply is more promising. According to Quong (2010, p. 308), we should simply ask "whether the particular act that is alleged to be protected by a right is consistent with the overall moral ideal which the system of rights is meant to uphold." Note that this view does not involve any perfectionist claims about the objective value of different autonomous conceptions of the good life (*pace* Raz 1991). Rather, this view delimits the scope of a right by

---

[3] To these we might add agents' interest in learning that others share their conceptions of the good, or similar ones (Raz 1991), although this is not strictly speaking a speaker-based interest.

insisting that each instance of the right's exercise be compatible with the overall mission of securing justice. Insofar as hate speech constitutes an attack on the ideal of justice—the ideal of treating others as free and equal citizens—it would not be protected (Quong 2010, p. 311; Waldron 1989, pp. 503–19, 518). Perhaps an adequate reply to this argument can be offered by those who seek to defend hate speech by appealing to the second moral power, but none has, in my view, yet been offered.

The Argument from Democracy

Arguably the most influential defense of freedom of speech appeals to our status as democratic citizens. While the details of the views vary, several authors have argued that we need free speech in order to engage in the enterprise of self-government—to author the very laws to which we are subject (Meiklejohn 1948, 1960; Sunstein 1993; Post 1991, 2011). On this view, public discourse, understood as discussion about political matters, is the paradigmatic kind of expression that the right to free speech serves to protect. Post (1991, p. 290) astutely identifies the essence of this family of views: "Democracy serves the value of self-determination by establishing a communicative structure within which the varying perspectives of individuals can be reconciled through reason."[5] I focus here on the most recent arguments in this family of views, developed by Corey Brettschneider and Eric Heinze (though see also the related view of Heyman 2009; cf. Simpson 2013).

Should hate speech be protected for the sake of democracy? The most developed recent arguments along these lines focus on hate speech in particular, holding that citizens of a democracy must be free to debate any viewpoint they see fit to debate, lest their society cease to be a democracy. Brettschneider, following Dworkin (2006, 2009) advances a view along these lines. "[I]t is essential to the legitimacy of…democracy," Brettschneider writes, "that [citizens] could choose to embrace inegalitarian principles and policies" (2012, p. 77). Heinze (2016, p. 5) elaborates a similar view, arguing that placing limits on the ideas that citizens are entitled to

[5] Baker (2009) defends a related view. While Post justifies free speech by appealing to democracy, Baker argues that both free speech and democracy are justified by the fundamental value of autonomy.

discuss in their deliberations over law and policy "encroaches upon the elements that *make* the state a democracy." Viewpoint neutrality—that fundamental commitment of US constitutional jurisprudence—is not merely a desirable feature of a democracy; it is a constitutive one.

Heinze's (2016) argument takes the form of a hypothetical imperative: If we wish to live in a democracy (as all of us purport to want), then we must endorse viewpoint neutrality and thus refuse to ban hate speech, lest we no longer live in a democracy. This raises the question of why we should want to live in a democracy, or at least a democracy of this demanding kind (Greene and Simpson 2017, pp. 757ff).[6] Perhaps Heinze simply seeks to be ecumenical, offering an argument that appeals to all democrats regardless of their reasons for being democrats. But this strategy is more limited than it seems, for even if we agree with Heinze's central stipulation—that society's democratic character is reduced when the law prevents citizens from advancing hateful policy proposals—it is not clear that this is a loss in anything valuable. That will depend on one's underlying normative democratic theory. Those who value democracy instrumentally, because of its tendency (epistemic or otherwise) to promote just outcomes (e.g., Arneson 2004, Estlund 2007, Landemore 2012) will welcome hate speech restrictions just in case they promote the achievement of those outcomes. Given that one potential rationale of hate speech restrictions is to stop the spread of the kind of xenophobic, nativist, hate-filled discourse that leads to flagrantly unjust policies, it is perfectly plausible that some instrumental democrats might welcome such restrictions.

But even those who value democracy noninstrumentally are not thereby fated to accept that democracy is objectionably diminished by laws banning hate speech. Suppose we value democracy because we think it is the system of government that best expresses respect for citizens as free and equal (Brettschneider 2007), or instantiates the public value of equality (Christiano 2008), or embodies a relational ideal of social equality (Kolodny 2014, Viehoff 2014), or constitutes a fair mode of decision making in the face of reasonable disagreement about justice (Waldron 1999), or constitutes a valuable relationship among moral agents (Beerbohm 2012). Holding these views does not commit one to the implausible view that every democratic decision has genuine value just by virtue of its status as a democratic decision; valuing

[6] For further critical engagement with Heinze's view and an opposing perspective, see Heinze and Phillipson 2018.

democracy non-instrumentally need not involve thinking that every democratic decision has value. After all, all of these views are compatible with the idea that the purpose of democracy—the purpose of government—is the achievement of substantive liberal justice (Howard 2018). Indeed, as Beerbohm (2012, p. 36) notes, the fact that a grave injustice is democratically authorized—the product of millions deciding together to do evil, rather than a lone despot—may aggravate its wrongness. Noninstrumental defenders of democracy, then, could concede that even if laws banning hate speech reduced the democratic character of their societies, these laws do not thereby reduce anything valuable. Democratic citizens need not be wronged by legislation that prevents them from publicly entertaining policies whose adoption they have no moral business even entertaining. Thus it is not clear why hate speech does not have the kind of immunity from restriction that Heinze (2016) suggests.

Unlike Heinze's, Brettschneider's (2012) argument takes the form of a categorical imperative, since he insists that democracy is morally required. Brettschneider appeals to Rawls's account of the two moral powers; because we have already discussed ideas related to the second moral power (involving the pursuit of the good life), let us focus on Brettschneider's appeal to what Rawls (2005, pp. 47ff) called the first moral power, the "capacity for a sense of justice." For Brettschneider (2012, p. 76), "[c]itizens must be free from coercive threat as they develop their own notion of justice." He concedes that while it is not "empirically necessary" for the development of the sense of justice that citizens be free to hear, promote, and embrace morally noxious views, "it would disrespect the independent judgment of free and equal citizens…if the state were to restrict their options." On his view, citizens must be free to accept or reject "the values of free and equal citizenship…rather than being coercively forced to do so" (Brettschneider 2012, p. 77).

However, bans on hate speech do not restrict people from having hateful thoughts; they simply prevent people from harming others through the expression of those thoughts. As I mentioned above, committed democrats need not think that citizens enjoy the authority to enact seriously unjust legislation of the sort that hate speech advocates. Brettschneider (2007) himself favors a system of strong judicial review, which ties the hands of democratic legislatures by preventing them from enacting legislation that undermines citizens' basic rights. Such a restriction, on Brettschneider's view, is not undemocratic, for it serves the very values that underwrite democracy. So if citizens may be permissibly prevented by a constitutional court

from enacting legislation that violates fundamental rights, why do they nevertheless have the prerogative to advocate the adoption of such hateful legislation in their public discourse? One fruitful future line of inquiry for those sympathetic to the democratic argument is to find a persuasive answer to this question.

The Thinker-Based Argument

The most recent argument for freedom of speech by a major scholar focuses neither on listener interests nor on speaker interests, but on the interests that we all have as thinkers. On Shiffrin's (2014, pp. 9–10) view, speech is special because it "provides the only precise mechanism by which one's mental contents may be conveyed to another mind, with all their subtlety and detail" (Shiffrin 2014, p. 10). Without knowledge of what others sincerely believe, our efforts to understand the world, pursue relationships, and cooperate on moral matters would verge on impossible. These concerns subsume all of the interests standardly invoked by free speech theorists, and more. Shiffrin's (2014, p. 92) argument thus connects to both Rawlsian moral powers. It is the value of open channels of linguistic communication that renders objectionable any attempt to subvert or restrict these channels, thereby obstructing human beings from telling one another what they genuinely believe.

I believe Shiffrin's striking position provides the most formidable argument yet in the literature for why hate speech ought to be permitted. Because her position does not value the transference of merely morally laudable thoughts, but rather the transference of thoughts *simpliciter*, it can explain clearly why hate speech should be protected. Because this position is a reasonably new one in the debate, the criticisms are still in progress, and no one has yet constructed an argument reconciling Shiffrin's view with bans on hate speech (though I am developing one in other work; see also Scanlon 2011). Can bans on hate speech be rendered consistent with Shiffrin's powerful theory? To my mind, this is yet another fruitful line of inquiry.

THE DUTY TO REFRAIN FROM HATE SPEECH

The first step in any argument for criminalizing hate speech is to establish that such speech is outside the protective ambit of the moral right to freedom of expression. The last section

conducted a critical review of the lively debate on precisely this proposition. Should the values that justify the moral right to free speech count in favor of protecting hate speech, then hate speech is, we can say, "morally protected" (Howard 2016, p. 32). And if it is morally protected, the fact that it has lamentable consequences—e.g., causing offense or inducing listeners to engage in violence and discrimination—is not a justification for censorship. In such cases, we should adopt the same attitude toward the hateful speaker as Mill [1978 (1859)], p. 9] enjoins us to adopt toward the agent poised to engage in self-harm; we have excellent reason for "remonstrating with him, or reasoning with him, or persuading him, or entreating him, but not for compelling him, or visiting him with any evil in case he do otherwise." Indeed, this reveals one rationale for the recurrent suggestion, inspired by Justice Louis Brandeis, that the right response to noxious speech is "more speech" [*Whitney v. California* 274 US 357 (1927)]; it is the only remedy that is morally available.

However, if the values that justify free speech do *not* count in favor of protecting hate speech, the case for criminalizing or otherwise restricting hate speech can proceed. Importantly, the mere fact that hate speech is morally unprotected is not sufficient to justify criminalization. Lots of activities—e.g., counting blades of grass—do not relate tightly to the weighty values that justify our most fundamental rights, what Rawls (2005, p. 294) called the basic liberties, and yet we do not therefore criminalize them. As the leading theorist of criminalization in the twentieth century put the point: "Liberty should be the norm; coercion always needs some special justification" (Feinberg 1984, p. 9). So even if hate speech is morally unprotected, that is only the beginning, not the end, of the argument for banning it.

To advance the argument, we must turn away from the question of rights and instead consider the matter of duties. Every criminal prohibition enforces putative duties to refrain from the prohibited conduct. To justify a criminal prohibition, we must show that agents have moral duties to refrain from the conduct in question (Husak 2008, p. 66; Tadros 2016b, p. 16). So the next step is to establish whether there is, in fact, an enforceable moral duty not to engage in hate speech, which bans on such expression serve to enforce.[7]

---

[7] One possibility is that there is, in fact, a moral duty to refrain from hate speech but that the function of the moral right to free speech is to render such a duty unenforceable (rather than simply foreclosing the duty's existence). I will set this subtle view aside, as the arguments for it

To that end, I review several possible arguments in defense of banning hate speech. These arguments are not typically presented in the scholarly literature as arguments about the duties of speakers not to engage in hate speech (cf. Post 1991, pp. 271–77, which instead focuses on harms). However, my suggestion is that they must be understood as such if they are to explain why hateful speakers are permissibly punished for their speech, and why the state does not wrong them when it threatens and enforces such punishment. For only if such coercion serves to enforce duties the speakers actually have can it be justified.

Why think that speakers have duties to refrain from hate speech? I will mention five, discussing the first two very briefly and the final three in some detail.

The Duty Not to Threaten

It is uncontroversial to suppose that directly threatening or intimidating another person, such that they reasonably fear for their safety, should constitute a criminal offense; such a principle is clearly embodied in the US Model Penal Code (§22) and upheld by the US Supreme Court [*Virginia v. Black* 538 US 343 (2003)]. Thus, expressions of hatred that take the form of direct, targeted intimidation are uncontroversially eligible for state interference. Many or even most instances of so-called targeted vilification fall into this category (Greenawalt 1995, p. 49; Brink 2001, p. 133; Yong 2011, pp. 394–96). This includes both face-to-face confrontations, especially those that deploy so-called fighting words, and other targeted actions, such as spray-painting a swastika on someone's house or leaving a noose hanging from their porch (Yong 2011, p. 394).

The Duty Not to Harass

It is a misdemeanor under the Model Penal Code (§240.4) to deliberately taunt, insult, or challenge another person "with purpose to harass". Several arguments against hate speech conceive it as a form of harassment (e.g., Nielsen 2004), pointing, for example, to the deleterious effects of racist harassment in particular, such as increased blood pressure and diminished self-esteem (Delgado & Stefanic 2018).

If the duty not to engage in hate speech derives from more general duties not to threaten or

---

would roughly be the same arguments that underlie freedom of speech canvassed in the previous section.

harass, hate speech bans may be redundant, as these wrongs are already typically illegal. But it is an interesting and important question whether the penalties for speech that threatens and harasses should be harsher if its content is hateful; this is simply one aspect of the debate over whether so-called hate crimes should face greater punishment than their nonhateful counterpart crimes (Wellman 2006). While that is controversial, the mere fact that there should be penalties for speech that explicitly or implicitly threatens harm is not. Accordingly, I set these two duties aside.

The Duty Not to Offend

Of course, not all "words that wound," to use Delgado's (1982) memorable phrase, do so by inspiring fear of subsequent attack by their speaker or by constituting harassment. That points us to a related but distinct third duty that is much more controversial, a duty not to offend (more precisely, a duty not to cause offense that cannot be reasonably avoided), where offense is defined as a kind of "unpleasant mental state" in response to the perception of wrongdoing Feinberg 1988, p. 2). Hate speech bans are often criticized on the grounds that we must permit people to offend others by expressing views or engaging in activity that shocks or upsets them (Waldron 1987). To show that causing offense is, in fact, properly forbidden would constitute a decisive reply to such a view.

The foremost defender of such a view is Feinberg (1988). Invoking a lineup of hypothetical lewd examples (in which public masturbation is among the tamest), he contends that "[i]t is always a good reason in support of a proposed criminal prohibition that it would probably be an effective way of preventing serious offense (as opposed to injury or harm) to persons other than the actor, and that it is probably a necessary means to that end"—principally, because the offense cannot be reasonably avoided by those who would be offended (Feinberg 1988, p. 2). Feinberg locates the wrong of hate speech within this paradigm, as it causes an objectionable form of psychic distress. For example, when neo-Nazis plotted their infamous 1977 march through a community of Holocaust survivors and other Jews in Skokie, Illinois, aiming to "insult them" and "lacerate their feelings," the wrong the neo-Nazis stood to perpetrate traced to its offensive

character (Feinberg 1988, p. 86).[8] Crucially, because the march was announced in advance and was therefore reasonably avoidable, Feinberg claims it would have been wrong to restrict it. Yet had it occurred spontaneously without warning, or had it become a frequent event, legal interference would have been warranted (Feinberg 1988, pp. 87–88); indeed, in such circumstances, offensive hate speech might also constitute the aforementioned wrong of harassment. Thus, on Feinberg's view, the permissibility of banning hate speech on offense-based grounds depends on how reasonable it is to expect those offended by it to avoid it. Such a view would condemn blanket bans on hate speech but permit restriction in certain circumstances.

Notwithstanding the intuitive force of Feinberg's examples of justifiably banned offensive acts (e.g., public bestiality or corpse desecration), the idea that there is an enforceable moral duty not to generate psychological distress in others—and that this is the right frame through which to think about hate speech—faces significant criticism. One criticism attacks Feinberg's view because it makes the wrongness of offense far too subjective. It is not enough, critics protest, that an act causes psychological distress; it must also be objectively wrongful—e.g., because it "treats others with a gross lack of consideration or respect"—such that it is fitting for it to generate that kind of psychological reaction (von Hirsch & Simester 2006, p. 120). Feinberg's (1988, p. 25) account of offense's normative significance is problematically content neutral; he is startlingly committed to the view that widespread disgust at interracial hand-holding in public, for example, is a genuine *pro tanto* reason to prohibit such conduct. While he concedes that this reason is clearly outweighed by countervailing considerations, we might sensibly ask why it receives any weight at all. Why does it even qualify as a moral reason? Indeed, in Millian spirit, arguably the discomfort one experiences in being offended in such a case may even have a positive moral valence, especially if it stirs one to think about and reconsider one's views (e.g., Waldron 1987). And some offense is simply inevitable, as in cases of deep religious disagreement (Waldron 2012, p. 131). Further, if one's negative psychological reaction to wrongful conduct traces to one's conviction that it is wrongful, then surely one would regard its wrongness as the justification for banning it, not its psychological effects. And if that is right—if the mere wrongness of disrespectful-but-not-harmful behavior (e.g., corpse desecration) is the proposed

[8] Post's (2009) powerful interpretation of hate speech bans as attempts to enforce social norms of civility or respect is a cousin of this view.

normative basis of proscribing that behavior—everything then hangs on the fraught question, much debated in legal philosophy, of whether only harmful wrongs may be criminalized or whether instead all (public) wrongs may be criminalized (Simester & von Hirsch 2011; cf. Duff 2014, Tadros 2016b**).**

The Duty Not to Defame

These pitfalls of the offense-based approach invite us to return to a harm-based approach to locating the wrong of hate speech. So let us consider a fourth possible duty that bans on hate speech serve to enforce. On Waldron's (2012, p. 39) view, the purpose of hate speech is to defame the members of historically marginalized groups—to degrade them as untrustworthy, inferior, lazy, dangerous, or unwanted—and as such to engage in a form of "group libel." While group libel is no longer recognized as an exception to the First Amendment, it briefly enjoyed recognition after the US Supreme Court decision *Beauharnais v. Illinois* (1952), which upheld a proscription against racist leaflets. Just as individual defamation undermines the target's ability to go through life in good standing, group defamation serves to make life in public unpleasant if not unbearable for its target. It does so by contaminating a vital public good, namely, "the public and visible assurance of just treatment that a society is supposed to provide to all of its members" (Waldron 2012, p. 81). Without this assurance, citizens live in constant fear that their basic rights, and the basic rights of their children, are not secure. They must constantly be alert for enemies in their midst who are prepared to violate those rights—both through threats of violence and discrimination and through attempts to enact unjust legislation that harms them.

But there is a genuine puzzle about this view. Waldron (2012, p. 169) insists that hate speech does not merely *cause* the diminution of the public good of assurance; it "constitutes the dispelling of the assurance". Simply by engaging in hate speech, and thereby communicating one's hatred to vulnerable citizens, the hatemongers instantaneously succeed in undermining assurance. Yet as one reviewer observed, "The harm is essentially caused by discovering what others think" (Alexander 2012). But this, he continues, "is surely a dubious basis for prohibiting hate speech. For the prohibition does not banish the attitude conveyed but only keeps its existence from being known. The ban may produce a sense of security, but it will be a false sense" (Alexander 2012). As Simpson (2013, p. 725) puts the same point:

> What we want is a social milieu in which people know that they will not be discriminated against, humiliated, or terrorised on account of their identity, because, as a matter of fact, it is *actually the case* they will not be discriminated against, humiliated, or terrorised on account of their identity.

Indeed, if anything, we might prefer to permit hate speech precisely in order to know who the hate-mongers are and what they think (Baker 2009; Barendt 2009, p. 453).

Waldron might reply that at least, by silencing the hatemongers, the state declares itself to be on the side of the vulnerable citizens and prepared to defend them.[9] The assurance that matters is not the false faith that all of your fellow citizens are liberal egalitarians but the conviction that your government has your back. Even so, one might wonder why the ordinary operations of the criminal law—punishing those who engage in violence, harassment, threats, and discrimination—is not sufficient to provide that assurance. If it did, the rationale for banning hate speech would need to be something other than the desire to avoid the revelation that citizens with hateful views exist.

The Duty Not to Incite Wrongdoing

That leads to a fifth and final possible duty violated by hateful speakers: the duty not to incite wrongdoing. Unlike the other duties, this duty forbids hate speech even when it is not communicated to the hated groups (e.g., when it is expressed on hateful websites that only likeminded people tend to visit; see Tsesis 2001). Hate speech that operates in this manner does not immediately cause harm, and it certainly does not constitute harm; rather, it operates by increasing the likelihood that some intervening agent, the listener, will engage in some kind of wrongdoing (see Schauer 1993 for this distinction). The most obvious category of wrongdoing here is criminal violence or unlawful discrimination (Brown 2008, 2015, pp. 66–71; Morgan 2007; Parekh 2012). But another category of wrongdoing is itself support for unjust legislation, such as "exclusionary or eliminationist policies" targeting vulnerable groups (Yong 2011, p. 398), or support for the politicians who champion such policies.

---

[9] This has affinities with the distinct view that the state has a duty to denounce hate speech to avoid complicity, defended by Brettschneider (2012, p. 71).

One way to deny the existence of such a duty is to insist that listeners are fully responsible for their own decisions and accordingly cannot lay blame at the feet of those who incite them. But this view is widely rejected (Nozick 1974, pp. 129–30; Amdur 1980, p. 294; Zimmerman 1985); clearly I can be blameworthy for the wrong of inciting you to commit violence, and you can simultaneously be blameworthy for failing to resist my exhortations. The fact that you are also to blame does not attenuate my blameworthiness as an inciter. Indeed, the fact that I have caused some wrong (say, the killing of an innocent person) in a way that involves your culpable agency may be worse than if I had just killed them myself, since it brings more wrongdoing into the world (Tadros 2016a, pp. 101–32; cf. Alexander 2000). Given the empirical claims on which this argument relies, about the causally contributory nature of hate speech to various forms of wrongdoing, this is a strong argument—to my mind, the strongest—for the existence of a duty to refrain from hate speech.[10]

ENFORCING DUTIES

Suppose that citizens lack a moral right to engage in hate speech, for it falls outside the protective ambit of the moral right to free speech. Suppose also that citizens have a moral duty to refrain from hate speech—a duty that is owed to others as a matter of public justice, correlating to others' rights, and so is, in principle, enforceable by the law. This does not, I stress, settle the question of whether the state is, all things considered, morally entitled to ban hate speech (of whatever specification is singled out by the relevant duty). There are all sorts of familiar reasons why we might deem it misguided, or even outright impermissible, to enforce this duty.

The first and most familiar class of reasons refers to the moral unreliability of the state. The idea is that we should not trust the state to enforce our moral duties with respect to speech, especially political speech. This view, well defended by Stone (2004) and others, is a powerful strain within the US constitutional tradition. There are (at least) three elements to it. The first

[10] For a skeptical view of the empirical claim, see Baker (2009) and Heinze (2016). For the argument that the causal relation is too tenuous to justify restriction, see Post (2009). For discussion of the idea that such speech should be banned as a precautionary measure in the face of empirical uncertainty, see Schauer (2009).

concerns our limited moral motivation. Politicians can be unscrupulous in the quest to maintain power, and there is an enduring worry that they will abuse statutes restricting speech to silence political adversaries. The second, related element concerns our limited epistemic competence about moral matters. Many of our fellow citizens hold mistaken views about what political morality requires and so are liable to suppress speech that they view as dangerous and objectionable but that is, in fact, perfectly legitimate. Viewpoint neutrality could be justified instrumentally as a bright-line rule that prevents them from enacting unjustified speech restrictions. Nor is it merely our political opponents whom we should regard as epistemically fallible. Even if we are certain we have grasped the moral truth on some matter, we should recognize that every previous generation has misguidedly insisted the same [Mill 1991 (1859)]. And while mistaken tax laws, for example, can be the subject of continuing discussion and revision, laws that restrict speech risk suppressing ongoing discussion about the very views silenced and whether we might have been mistaken to silence them. Finally, there are inherent difficulties in authoring statutes that are not objectionably over-inclusive, leading to the prosecution of speakers who do not merit prosecution. This makes us unavoidably and heavily reliant on prosecutorial discretion in the enforcement of hate speech restrictions, which compounds the previous worries.

Even if the state could be trusted to restrict hate speech in a morally responsible fashion—something the European experience with hate speech laws suggests is at least possible—that fact would still not settle the case for restrictions. For even if statutes were authored impeccably, they may nevertheless be counterproductive, for at least two reasons. First, bans on hate speech might antagonize hateful citizens and deprive them of a valuably cathartic legal outlet for their rage (e.g., Rosenblum 1998, p. 254ff; see also Emerson 1970, Strossen 1990). Second, there may be a better mechanism for combating hate than the criminal law: in the words of Brandeis, “more speech.” On this view, rather than suppress hateful ideas, the best response is to argue against them (Sunstein 2018, p. 248).

Counter-speech tends to be defended by those who think that the moral right to freedom of expression protects hate speech (e.g., Brettschneider 2012), or who think that the state is morally unreliable. This makes sense; if banning hate speech is morally misguided (for either reason), counter-speech is the only morally permissible remedy. Yet even if we view hate speech as morally unprotected, and even if we would generally trust the state to ban it, it could still be the

case that counter-speech is simply more effective at combating hateful views than is the law. That would itself be a powerful consideration against criminalization.

There is a small but burgeoning literature, largely among practitioners, about the best methods of counter-speech (e.g., Benesch 2012; Benesch et al. 2016; Brown 2016). But at present, the topic of counter-speech in normative political theory—who should do what, and why—is seriously underdeveloped. There are, to be sure, welcome exceptions.[11] Brettschneider (2012), for example, holds that the state itself must argue against hate speech, both through its official communications and through other expressive decisions it makes (e.g., concerning what organizations receive tax-exempt status). By doing so, the state endeavors to persuade citizens with hateful views "to adopt the values of equal citizenship" (Brettschneider 2012, p. 72). Brettschneider's account provides the foundation for a promising research agenda regarding state speech, but it leaves almost entirely open what it is that ordinary citizens should be expected to do to combat hate in their daily lives. This is a vital question. Despite the power of the state, its persuasive potential has limits. As Brettschneider recognizes, while the state can try to persuade hateful citizens by reiterating its commitment to fundamental liberal values, it cannot take a deeper stand on why those values are correct without endorsing a particular comprehensive philosophical doctrine—something the state, according to prominent political liberals, is not supposed to do (Rawls 2005, p. 140; Brettschneider 2012, p. 105). Yet it seems perfectly clear that while the state may be forbidden to engage in such comprehensive persuasion efforts, ordinary citizens are permitted to, perhaps even required to. Happily, normative theorists are beginning to attend to this question (e.g., Schwartzman 2012, Clayton & Stevens 2014**,** Badano & Nuti 2018). Future normative scholarship on free speech would do well to focus on the issue of counter-speech, both offline and online.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

[11] Arguably, work on rhetoric in political theory falls into this category (see Garsten 2009, 2011; Chambers 2009).

ACKNOWLEDGMENTS

In am indebted to Joe Atkinson for invaluable research assistance and philosophical discussion. I am grateful to the British Academy and the Leverhulme Trust for research funding. Thanks also to Amanda Greene and Robert Simpson for organizing a symposium in London on Eric Heinze's book, *Hate Speech and Democratic Citizenship*—and to Eric himself for spirited replies on that occasion, which enabled me to think about the democratic argument for protecting hate speech. Thanks, too, to Matteo Bonotti and Jonathan Seglow for organizing a workshop in London on hate speech law, and to Eric Barendt, on whose paper I commented at that event. This article, alongside my other work on free speech, has also benefitted from comments and feedback from Christian Barry, Corey Brettschneider, Bob Goodin, Malte Ibsen, Cécile Laborde, Emily McTernan, Shmulik Nili, Tom Parr, Avia Pasternak, Gavin Phillipson, Jiewuh Song, Laura Valentini, Han van Wietmarschen, and Jeremy Waldron. While I never presented this article in its present form, elements of the analysis appear in other work I have presented, and so it indirectly has benefitted from audiences at Nuffield College, Oxford; the Australian National University; the University of Canberra; the American Philosophical Association (Pacific Division); the London School of Economics; and UCL.

LITERATURE CITED

Alexander L. 2000. Incitement and freedom of speech. In *Freedom of Speech and Incitement Against Democracy*, ed. D Kretzmer, F Hazan, pp. 101-118. The Hague: Kluwer Law

Alexander L. 2005. *Is There a Right to Freedom of Expression*? Cambridge, UK: Cambridge Univ. Press

Alexander L. 2012. Fighting for words. *Natl. Rev.* Dec. 17. https://www.nationalreview.com/magazine/2012/12/17/fighting-words/

Amdur R. 1980. Scanlon on freedom of expression. *Philos. Public Aff.* 9(3):287–300

Arneson R. 2004. Democracy is not intrinsically just. In *Justice and Democracy: Essays for Brian Barry*, ed. K Dowding, R Goodin, C Pateman, pp. 40-58. Cambridge, UK: Cambridge Univ. Press

Badano G, Nuti A. 2018. Under pressure: political liberalism, the rise of unreasonableness, and

the complexity of containment. *J. Political Philos.* 26(2):145–68

Baker CE. 1978. Scope of the first amendment freedom of speech. *UCLA Law Rev*. 25:964–1040

Baker, CE. 1997. Harm, liberty, and free speech . S. California Law Rev. 70:979-1020.

Baker CE. 2009. Autonomy and hate speech. See Weinstein & Hare 2009, pp. 139–57

Bambauer D. 2006. Shopping badly: cognitive biases, communications, and the fallacy of the marketplace of ideas. *Univ. Colorado Law Rev.* 77:649–710

Barendt E. 2009. Incitement to, and glorification of, terrorism. See Weinstein & Hare 2009, pp. 445-462

Beerbohm E. 2012. *In Our Name: The Ethics of Democracy*. Princeton, NJ: Princeton Univ. Press

Benesch S. 2012. *Dangerous speech: a proposal to prevent group violence*. Jan. 12, World Policy Institute. Report available at https://worldpolicy.org/wp-content/uploads/2016/01/Dangerous-Speech-Guidelines-Benesch-January-2012.pdf

Benesch S, Ruths D, Dillon KP, Saleem HM, Wright L. 2016. *Considerations for successful counterspeech*. Dangerous Speech Proj., Report available at https://dangerousspeech.org/considerations-for-successful-counterspeech.

Brettschneider C. 2007. *Democratic Rights: The Substance of Self-Government*. Princeton, NJ: Princeton Univ. Press

Brettschneider C. 2012. *When the State Speaks, What Should It Say? How Democracies Can Protect Expression and Promote Equality*. Princeton, NJ: Princeton Univ. Press

Brietzke P. 1997. How and why the marketplace of ideas fails. *Valparaiso Univ. Law Rev.* 31:951–69

Brink, D. O. 2001. Millian principles, freedom of expression and hate speech. *Legal Theory* 7:119–157.

Brison S. 1998. The autonomy defense of free speech. *Ethics* 108(2):312–39

Brown A. 2008. The Racial and Religious Hatred Act 2006: a Millian response. *Crit. Rev. Int. Soc. Political Philos.* 11(1):1–24

Brown A. 2015. *Hate Speech Law: A Philosophical Examination*. London: Routledge

Brown A. 2017a. What is hate speech? Part 1: The myth of hate. *Law Philos.* 36(4):419–68

Brown A. 2017b. What is hate speech? Part 2: Family resemblances. *Law Philos.* 36(5):561–613

Brown A. 2017c. The “who?” question in the hate speech debate. Part 1: Consistency, practical,

and formal approaches. *Can. J. Law Jurisprudence* 29(2):275–320

Brown A. 2017d. The "who?" question in the hate speech debate. Part 2: Functional democratic approaches. *Can. J. Law Jurisprudence* 30(1):23–55

Brown R. 2016. Defusing hate: a strategic communication guide to counteract dangerous speech. Washington, DC: US Holocaust Memorial Museum. Report available at https://www.ushmm.org/confront-genocide/how-to-prevent-genocide/hate-speech-and-incitement-to-genocide/defusing-hate-a-guide-to-counteract-dangerous-speech)

Chambers S. 2009. Rhetoric and the public sphere: Has deliberative democracy abandoned mass democracy? *Political Theory* 37:323–50

Christiano T. 2008. *The Constitution of Equality*. Oxford, UK: Oxford Univ. Press

Clayton M, Stevens D. 2014. When God commands disobedience: political liberalism and unreasonable religions. *Res. Publ.* 20:65–84

Cohen J. 1993. Freedom of expression. *Philos. Public Aff.* 22(3):207–63

Darwall S. 1977. Two kinds of respect. *Ethics* 88(1):36–49

Delgado R. 1982. Words that wound: a tort action for racial insults, epithets, and name-calling. *Harvard Civil Rights–Civil Liberties Law Rev*. 17:133-181

Delgado R, Stefanic J. 2018. *Must We Defend Nazis? Why the First Amendment Should Not Protect Hate Speech and White Supremacy*. New York: NYU Press

Duff RA. 2014. Towards a modest legal moralism. *Criminal Law Philos*. 8(1):217–35

Dworkin R. 1996. *Freedom's Law: The Moral Reading of the American Constitution*. Oxford, UK: Oxford Univ. Press

Dworkin R. 2006. A new map of censorship. *Index Censorship* 35(1):130–33

Dworkin R. 2009. Forward. See Weinstein & Hare 2009, pp. 123–38

Emerson T. 1970. *The System of Freedom of Expression*. New York: Vintage

Estlund D. 2007. *Democratic Authority*. Princeton, NJ: Princeton Univ. Press

Feinberg J. 1984. *The Moral Limits of the Criminal Law, Volume 1: Harm to Others*. Oxford, UK: Oxford Univ. Press

Feinberg J. 1988. *The Moral Limits of the Criminal Law, Volume 2: Offense to Others*. Oxford, UK: Oxford Univ. Press

Fish S. 1994. *There's No Such Thing as Free Speech…and It's a Good Thing Too*. Oxford, UK: Oxford Univ. Press

Garsten B. 2009. *Saving Persuasion: A Defense of Rhetoric and Judgment*. Cambridge, MA: Harvard Univ. Press

Garsten B. 2011. The rhetoric revival in political theory. *Annu. Rev. Political Sci.* 14:164–74

Garton Ash T. 2016. *Free Speech*. London: Atlantic Books

Glaeser E, Sunstein C. 2014. Does more speech correct falsehoods? *J. Legal Stud.* 43(1):65–93

Gordon J. 1997. John Stuart Mill and the 'marketplace of ideas'. *Soc. Theory Pract.* 23(2):235–49

Greenawalt K. 1989. *Speech, Crime, and the Uses of Language*. Oxford, UK: Oxford Univ. Press

Greenawalt K. 1995. Fighting Words: Individuals, Communities and Liberties of Speech. Princeton, NJ: Princeton Univ. Press.

Heinze E. 2016. *Hate Speech and Democratic Citizenship*. Oxford, UK: Oxford Univ. Press

Heinze E, Phillipson G. 2018. *Debating Hate Speech*. Oxford, UK: Hart Publishing.

Heyman S. 2009. Hate speech, public discourse, and the first amendment. See Weinstein & Hare 2009, pp. 123–38

Howard JW. 2016. Moral subversion and structural entrapment. *J. Political Philos.* 24:24–46

Howard JW. 2018. The labors of justice: democracy, respect, and judicial review. *Crit. Rev. Int. Soc. Political Philos.* In press. https://doi.org/10.1080/13698230.2017.1334437

Husak D. 2008. *Overcriminalization: The Limits of the Criminal Law*. Oxford, UK: Oxford Univ. Press

Ingber S. 1984. The marketplace of ideas: a legitimizing myth. *Duke Law J.* 1984(1):1–91.

Kendrick L. 2017. Free speech as a special right. *Philos. Public Aff.* 45(2):87–117

Kolodny N. 2014. Rule over none. II: Social equality and the justification of democracy. *Philos. Public Aff.* 42(4):287–336

Landemore H. 2012. *Democratic Reason*. Princeton, NJ: Princeton Univ. Press

MacKinnon CA. 1996. *Only Words*. Cambridge, MA: Harvard Univ. Press

Marshall W. 1995. In defense of the search for truth as a first amendment justification. *Ga. Law Rev.* 30:1–39

Matsuda M. 1989. Public response to racist speech: considering the victim's story. *Michigan Law Rev.* 87:2320-2381.

Meiklejohn A. 1948. *Free Speech and Its Relation to Self-Government*. New York: Harper and Brothers

Meiklejohn A. 1960. *Political Freedom*. New York: Harper and Brothers

Mill JS. 1978 (1859). *On Liberty*, ed. E Rapaport. Indianapolis, IN: Hackett

Morgan G. 2007. Mill's liberalism, security, and group defamation. In *Freedom of Expression: Counting the Costs*, ed. G Newey, pp. 121-143. Cambridge, UK: Cambridge Scholars Press

Nagel T. 2002. *Concealment and Exposure*. Oxford, UK: Oxford Univ. Press

Nielsen L. 2004. *License to Harass: Law, Hierarchy, and Offensive Public Speech*. Princeton, NJ: Princeton Univ. Press

Nozick R. 1974. *Anarchy, State, and Utopia*. New York: Basic Books.

Parekh B. 2012. Is there a case for banning hate speech? In *The Content and Context of Hate Speech: Rethinking Regulation and Responses*, ed. M Herz, P Molnar, pp. 37–56. Cambridge, UK: Cambridge Univ. Press

Post R. 1991. Racist speech, democracy, and the first amendment. *William Mary Law Rev.* 32:267–327

Post R. 2009. Hate speech. See Weinstein & Hare 2009, pp. 123–38

Post R. 2011. Participatory democracy as a theory of free speech: a reply. *Virginia Law Rev.* 97:617–32

Quong J. 2010. *Liberalism Without Perfection*. Oxford, UK: Oxford Univ. Press

Rawls J. 2005. *Political Liberalism*. New York: Columbia Univ. Press

Raz J. 1991. Free expression and personal identification. *Oxford J. Legal Stud.* 11(3):303–24

Rosenblum N. 1998. *Membership and Morals: The Personal Uses of Pluralism in America*. Princeton, NJ: Princeton Univ. Press

Scanlon T. 1972. A theory of freedom of expression. *Philos. Public Aff.* 1(2):204–26

Scanlon T. 1979. Freedom of expression and categories of expression. *Univ. Pittsburgh Law Rev.* 40:519–50

Scanlon T. 2011. Comment on Shiffrin's thinker-based approach to freedom of speech. *Const. Comment.* 27:327–35

Schauer F. 1982. *Free Speech: A Philosophical Inquiry*. Cambridge, UK: Cambridge Univ. Press

Schauer F. 1983. Must speech be special? *Northwest. Univ. Law Rev.* 78:1284

Schauer F. 1993. The phenomenology of speech and harm. *Ethics* 103(4):6350–653

Schauer F. 2009. Is it better to be safe than sorry? Free speech and the precautionary principle. *Pepperdine Law Rev.* 36(2):301–16

Schauer F. 2010. Facts and the first amendment. *UCLA Law Rev*. 57:897–919

Schauer F. 2015. Free speech on Tuesdays. *Law Philos*. 34:119–40

Schwartzman M. 2012. The ethics of reasoning from conjecture. *J. Moral Philos.* 9(4):521–44

Scoccia D. 1996. Can liberals support a ban on violent pornography? *Ethics* 106(4):776–99

Shiffrin S. 2003. Speech, death, and double effect. *New York Univ. Law Rev.* 78:1135–85

Shiffrin S. 2014. *Speech Matters*. Princeton, NJ: Princeton Univ. Press

Simester A, von Hirsch A. 2011. *Crimes, Harms, and Wrongs*: *On the Principles of Criminalisation*. Oxford, UK: Hart Publishing

Simpson R. 2013. Dignity, harm, and hate speech. *Law and Philosophy* 32(6):701-28.

Sparrow R, Goodin R. 2007. The competition of ideas: market or garden? *Crit. Rev. Int. Soc. Political Philos.* 4(2):45–58

Stone G. 1987. Content-neutral restriction. *Univ. Chicago Law Rev.* 54:46–118

Stone G. 1994. Hate speech and the U.S. Constitution. *East Eur. Const. Rev.* 3:78–82

Stone G. 2004. *Perilous Times*. New York: W.W. Norton

Strossen N. 1990. Regulating racist speech on campus: a modest proposal? *Duke Law J*. 39:484–572

Sunstein C. 1986. Pornography and the first amendment. *Duke Law J*. 34(4):589–627

Sunstein C. 1993. *Democracy and the Problem of Free Speech*. New York: Free Press

Sunstein C. 2018. *#Republic: Divided Democracy in the Age of Social Media*. Princeton, NJ: Princeton Univ. Press

Tadros V. 2016a. Permissibility in a world of wrongdoing. *Philos. Public Aff.* 44(2):101–32

Tadros V. 2016b. *Wrongs and Crimes*. Oxford, UK: Oxford Univ. Press

Tsesis A. 2001. Hate in cyberspace: regulating hate speech on the internet. *San Diego Law Rev.* 38:817

Waldron J. 1987. Mill and the value of moral distress. *Political Stud*. 35(5):410–23

Waldron J. 1989. Rights in conflict. *Ethics* 99(3):503–19

Waldron J. 1999. *Law and Disagreement*. Oxford, UK: Oxford Univ. Press

Waldron J. 2012. *The Harm in Hate Speech*. Cambridge, MA: Harvard Univ. Press

Weinstein J, Hare I, eds. 2009. *Extreme Speech and Democracy*. Oxford, UK: Oxford Univ. Press

Wellman C. 2006. A defense of stiffer penalties for hate crimes. *Hypatia* 21(2):62–80

Viehoff D. 2014. Democratic equality and political authority. *Philos. Public Aff.* 42(4):337–75

Volokh E. 2011. In defense of the marketplace of ideas/search for truth as a theory of free speech protection. *Va. Law Rev.* 97:595–601

von Hirsch A, Simester A. 2006. Penalising offensive behaviour: constitutive and mediating principles. In *Incivilities: Regulating Offensive Behaviour*, ed. A von Hirsch, A Simester, pp. 115-132. Oxford, UK: Hart Publishing

Yong, Caleb. 2011. Does freedom of speech include hate speech? *Res Publica*. 17:385-403.

Zimmerman M. 1985. Intervening agents and moral responsibility. *Philos. Q.* 35(141):347–57