



Introduction: Hate, Offence and Free Speech in a Changing World

Paul Billingham¹ · Matteo Bonotti²

Accepted: 6 August 2019 / Published online: 5 September 2019
© Springer Nature B.V. 2019

1 Current Debates on Hate and Offensive Speech

The recent Brexit referendum in the UK, Donald Trump's election as President of the United States, the rise of populism and far-right politics in several European countries, and the current wave of Islamophobia across Europe, prompted by migratory pressure and an unstable Middle East, have brought hate speech to the forefront of both public and academic debate. Offensive speech, especially that directed at members of religious minorities, also continues to elicit debate, as shown by the 2006 *Jyllands-Posten* Muhammad cartoons controversy and, more recently, by the *Charlie Hebdo* controversies and attacks.

Political and legal theorists have for a long time been concerned with the normative dimensions of free speech. Earlier works tended to focus on the very justification for freedom of speech, ranging from John Stuart Mill's argument from truth (2006) to arguments grounded in the value of individual autonomy (Scanlon 1972, 1979; Dworkin 1981; Baker 1989) and democracy (Meiklejohn 1961; Sunstein 1993). These earlier reflections of the foundations of free speech were soon followed, especially in the 1980s and 1990s, by the rise of an extensive literature on pornography and free speech, with many feminist thinkers challenging the general liberal presumption in favour of free speech and endorsing anti-pornography censorship interventions (MacKinnon 1993), sometimes drawing on the very liberal foundations that their opponents relied upon (Dyzenhaus 1992; Langton 1990).

While the debate concerning pornography continues to elicit interest in both political and legal theory, attention over the past twenty years has rapidly shifted towards hate and offensive speech, especially speech targeting members of religious and ethnic minorities. The wave of Islamophobia that followed 9/11 in the USA and beyond made it urgent for scholars to reflect on, and indeed question, longstanding assumptions about the value of unconstrained freedom of speech, especially as countries such as the UK introduced stricter regulations on religious hatred.

✉ Matteo Bonotti
Matteo.Bonotti@monash.edu

¹ Department of Politics and International Relations, University of Oxford, Oxford, UK

² Department of Politics and International Relations, School of Social Sciences, Monash University, Melbourne, Australia

Scholars examining hate speech have done so from a variety of perspectives. Some, for example, have embraced a Millian consequentialist approach and argued that hate speech, especially incitement to hatred, may permissibly be regulated if we can show that it can harm others by contributing to a climate of hatred leading to discrimination, violence and injustice against members of certain social groups (Brown 2008). This kind of direct consequentialist approach, however, heavily depends on empirical evidence regarding the effects of hate speech that may be uncertain and/or difficult to ascertain. Therefore some authors have argued that to understand what is morally wrong with hate speech we need to move beyond the harm principle and draw on an ‘offense principle’ (Feinberg 1985). Under this conception, it has been argued (van Mill 2018), hate speech belongs to the category of ‘profound offenses’, which occur when ‘one is outraged at the offending conduct on grounds quite independent of its effect on oneself’ (Feinberg 1985, p. 93). Still others have drawn on different aspects of Mill’s theory, stressing the importance of ‘high-value’ speech in his argument from truth (Brink 2008). Unlike ‘low-value’ speech, ‘high-value’ speech engages our deliberative capacities and therefore deserves protection. But many instances of hate speech are ‘low-value’, for various reasons. They might not engage our deliberative faculties at all. For example, ‘fighting words’, which Brown (2015, p. 36) includes among ‘expression-oriented hate crimes’, belong to this category (Brink 2008) and are indeed subject to regulation even under such a permissive free speech regime as the one operative in the USA. But even if some types of hate speech do engage our deliberative faculties, they make the discovery of truth (which is central to Mill’s account) very difficult (Brown 2015, p. 114). They also might prevent some individuals (i.e. the targets of hate speech) from participating in that very process of truth discovery, a point often highlighted in campus speech codes at various universities across the Western world.

Partly due to the problems affecting the Millian approach, other authors have preferred to move towards a more indirect consequentialist approach that focuses on the ‘harmful albeit diffusive effects’ (Heinze 2016, p. 137) of hate speech. Especially influential in this area has been the work of Jeremy Waldron (2012) who, drawing more or less explicitly on authors such as Catherine MacKinnon (1993) and Rae Langton (1993), argues that hate speech undermines individuals’ assurance of their equal public standing and thereby injures their civic dignity.

In response to these arguments, critics of hate speech laws have claimed that such laws undermine such key values as individual autonomy (Baker 1989), democratic citizenship (Heinze 2016) and democratic legitimacy (Dworkin 2009). Alexander Brown (2015) has provided an especially thorough and insightful analysis of hate speech laws, by disaggregating such laws into different clusters, examining arguments for and against them, and developing criteria for principled compromise between those arguments.

Debates on offensive speech, and especially religiously offensive speech, have an even longer pedigree in Western liberal democracies, perhaps because many had blasphemy laws that were used to protect the established religion(s) from offensive attacks and criticisms. Indeed, such a law existed in the UK until 2008. No one was prosecuted under this law between 1922 and 1977, but in 1977 the newspaper *Gay News* was convicted due to the publication of James Kirkup’s poem ‘The Love That Dares to Speak Its Name’ (1976), which included a graphic account of sexual intercourse between Jesus and a Roman centurion and was therefore considered blasphemous. More recently, the publication of Salman Rushdie’s *The Satanic Verses* (1988), which contained a portrayal of Islam that many Muslims considered blasphemous, elicited a lively debate among political theorists, with some criticizing (Jones 1990a, 1990b) and others endorsing (Parekh 2006) regulation of this kind of speech. An attempt to prosecute Rushdie under the UK blasphemy law failed, as did a similar attempt to

prosecute BBC director general Mark Thompson for the broadcast of the musical *Jerry Springer: The Opera* (2003), which contained depictions of God, Jesus and the Virgin Mary that many viewers considered blasphemous. Both failed prosecutions revealed important normative challenges. The Rushdie case highlighted that UK blasphemy law only protected Christianity, while leaving non-Christian religious and non-religious beliefs subject to offence. This was inconsistent with basic principles of liberal neutrality, and was one of the reasons for the law's eventual repeal. In the Thompson case, the case for free speech (and against prosecution) was strengthened by the UK 1968 Theatres Act and 1990 Broadcasting Act, which provide freedom of expression in theatrical and broadcasting productions with a special level of protection, thus raising the question of whether liberal states should value and protect certain forms of speech more than others.

As in the case of hate speech, arguments for the regulation of blasphemous speech are diverse. Some have argued that anti-blasphemy laws protect 'the fabric of society' (Warburton 2009, p. 44), although this might seem like an anachronistic argument in increasingly multicultural societies, and one that would seem to be mainly aimed at preserving 'the social and political status quo' (Hare 2009, p. 301). Another argument is that religion and 'a sense of the sacred' (Barendt 2005, 187) are intrinsically valuable and should be protected through state laws. Or, in a way that seems to reflect Feinberg's (1985) aforementioned 'offense principle', one might argue that 'it is right for the law to protect Christians [or followers of any religion] against the shock and outrage to their feelings provoked by the publication of blasphemous material' (Barendt 2005, p. 188).

Especially rich has been the debate on offensive speech sparked by the publication of 12 cartoons of the prophet Muhammad by the Danish newspaper *Jyllands-Posten* in September 2005. Some of the arguments made in that debate have relevance well beyond the specific case examined. Robert Post (2007), for example, objects to regulation by appealing to the importance of freedom of speech for public discourse and democratic legitimacy. Sune Lægård (2007) argues instead that the main arguments normally provided by those criticizing the cartoons (including offence to religious sensibilities, and the view that the cartoons undermined Muslims' identity and self-respect and contributed to their oppression within society) were not sufficiently strong to warrant their regulation. Christian F Rostbøll (2009) invokes a moral duty of self-regulation grounded in the ideal of individual autonomy. More specifically, Rostbøll claims, a Millian conception of autonomy, in which the critical exercise of our deliberative faculties is central, may lead to an arrogant and disrespectful approach to other people's beliefs. We should embrace instead a Kantian view of autonomy, according to which others should be respected as persons 'the realization of whose ends is as important as the realization of our own' (Rostbøll 2009, p. 633). This respect should be manifested through self-censorship, even in the absence of laws regulating offensive speech.

2 Overview of the Papers in the Special Issue

The articles in this special issue provide fresh insights into the broad debates illustrated in the previous section, by critically assessing and expanding key theoretical perspectives and applying them to novel social and political issues.

One of the most prominent recent contributions to the debate on hate speech, as we have seen, is Waldron's *The Harm in Hate Speech* (2012). Waldron argues that hate speech undermines individuals' assurance of their equal public standing and thereby injures their dignity. This claim is the focus of Eric Barendt's contribution to this special issue. Barendt

argues that it is unclear whether Waldron believes that hate speech causes harm or constitutes harm. This matters because the arguments required to justify these two claims have importantly different structures and implications. Careful analysis leads Barendt to conclude that Waldron is best understood as making a ‘weak consequentialist’ argument – an argument that hate speech tends to cause harm. The case for prohibition thus depends on the empirical evidence for this causal claim, and would need to weigh this likelihood of harm against the disvalue of restrictions on speech. The constitutive argument, on the other hand, does not require this kind of empirical evidence, and it might imply that we should place hate speech into the constitutional category of ‘conduct’, rather than the more protected category of ‘speech’. Barendt suggests that Waldron also makes this constitutive claim, but criticises this argument as unsupported by the Austinian speech-act theory on which it is meant to rely. Waldron’s case for legal bans on hate speech thus depends on the force of his weak consequentialist argument and is not strengthened by the constitutive argument.

Several other theorists, most notably Rae Langton (1993), have drawn on speech-act theory to argue that hate speech does indeed constitute harm, since it is illocutionarily subordinating. Barendt rejects such arguments on the basis that hate speakers normally lack the kinds of authority that would be required for their speech to have this illocutionary force. In her contribution, however, Suzanne Whitten argues that questions regarding the authority of hate speakers in fact open up the possibility that targets of hate speech can themselves influence, and indeed mitigate, some of the harms that they might suffer, by rejecting the authority claims of the speaker. We can best appreciate this insight, Whitten argues, by adopting an intersubjective, recognition-sensitive approach, which focuses on the dynamics of recognition and misrecognition in specific instances of hate speech. Whitten illustrates this with a real-life example in which an instance of hate speech might have constituted harm by entrenching racist norms, but did not damage the target’s sense of self, since the target rejected the speaker’s claims to misrecognise her, by refusing to recognise his claimed authority. The upshot of Whitten’s discussion is that we can reduce harmful misrecognition by creating an environment where minorities are empowered to speak back against hate speakers and refuse to recognise them as having authority. Whitten suggests that this could be done through the promotion and resourcing of counter-narratives.

Wrongful speech comes in many forms, and several of the contributions to this special issue seek to distinguish different kinds. This is important because different policies and strategies might be justifiable and/or effective in response to different kinds of speech. In her contribution, Louise Richardson-Self identifies and analyses five categories of ‘cis-hetero-misogynistic speech’ found in the Facebook comments section of a prominent Australian newspaper. Richardson-Self argues that there has been an evolution of hetero-misogynistic speech, from calling women lesbians and masculine to explicitly asserting that they are men, or not female, or even that they are androgynous ‘its’. Importantly, while this speech has emerged in response to transgender and queer movements, it is often targeted at cisgender, heterosexual women. It is thus not simply anti-queer speech, but misogyny, aimed at women who are perceived as in some way challenging the patriarchal gender order, with the aim of reasserting and reinforcing patriarchal norms and behavioural expectations. While this speech disincentivises allyship between cisgender and queer women, it also shows that promoting justice for the latter is in the interests of all women. Richardson-Self thus offers some reflections on what could be done about the categories of wrongful speech that she identifies, in terms of both prohibiting it and (more fundamentally) bringing about an egalitarian gender order that will reduce the occurrence of such speech.

While Richardson-Self still considers the speech that she examines to be hate speech, Matteo Bonotti and Jonathan Seglow expand the debate by analysing when religiously offensive speech – speech that directly targets beliefs, not believers – might be wrongful. Many theorists distinguish hate speech from merely offensive speech and argue that only the former is apt for regulation on grounds of wrongfulness. Bonotti and Seglow disagree, arguing that, against a background of structural injustice, religiously offensive speech can undermine its targets' self-respect and non-domination, and thus be wrongful. Such speech reduces members of minority religious groups to de facto second-class citizens, impairing their equal civic status. The wrongfulness of this speech gives at least a *pro tanto* case for regulation. Bonotti and Seglow conclude by considering the strengths and weaknesses of four possible non-censorious responses to such speech that might be used if regulation is warranted all-things-considered.

Katharine Gelber also distinguishes hate speech from a different category of speech, namely anti-Western terrorist-extremist speech. The latter form of speech is often elided with hate speech, but Gelber argues that it is in fact importantly different in both its audience and mechanisms of harm. Both types of speech have a direct audience of people who might be persuaded to adopt the worldview being promulgated, but only hate speech also has as its audience minorities who are targets of discursive harm by virtue of their membership of a marginalised group. In this way, hate speech constitutively harms individuals identified on the basis of minority group membership. Anti-Western terrorist-extremist speech, in contrast, can harm any member of society, and does so causally, through creating generalised fear that one might become the victim of a terrorist attack. The differences between these types of speech are more significant than their similarities, such that different policy tools are needed in relation to each in order to mitigate or address their harms.

As well as being of different kinds, wrongful speech can also arise in different contexts. Andrew Reid's article focuses on hateful speech used within political campaigns. Reid argues that those who use such speech might justifiably be subject to non-criminal sanctions, such as the withdrawal of institutional support and recognition. This is the case even if other kinds of misleading campaign material, such as material making inaccurate factual claims, has a greater effect on actual voting behaviour. Inaccurate material can be challenged by citizens, whereas hateful material creates an environment where some individuals cannot participate fully in politics. This latter situation can undermine democratic legitimacy, according to Reid's conception of the effective equal political voice that is required by the ideal of democratic public justification. Regulation that protects individuals' equal democratic status can be justified even if it infringes on other speakers' formal democratic rights. Reid illustrates and supports his argument by analysing two pieces of prominent campaign material used during the Brexit referendum campaign.

Several of the articles discussed thus far mention the possibility of non-coercive state responses to wrongful speech, such as counter-speech and withdrawing funding and benefits. Paul Billingham's contribution critically assesses a version of this 'transformative liberal' approach developed by Corey Brettschneider (2012). According to Brettschneider, the state should engage in 'democratic persuasion', by actively but non-coercively opposing groups whose views or practices are hostile to, or implausibly compatible with, the central liberal democratic value of free and equal citizenship. Billingham presents a series of critiques of Brettschneider's account, arguing that two of its central justifications are implausible, some of Brettschneider's favoured policies are in fact coercive, his account requires the state to make complex and contestable judgments of a kind that it ought not make, his view has various

troubling implications, and it ultimately rests on an overly-thick conception of free and equal citizenship and thus an overly-broad definition of hateful viewpoints. While democratic persuasion might be justified against truly hateful groups, Brettschneider's view is objectionably far-reaching.

In keeping with the literature, most of the articles in this special issue focus on ways in which the state might tackle hate speech and other forms of wrongful speech. The final contribution, by Simon Thompson, takes a different approach. Thompson considers the reasons that individuals might themselves have to freely choose not to speak hatefully of others, even when they have a desire to do so. In other words, he explores possible reasons for self-restraint on the part of prospective hate speakers. Thompson notes that for intentional hate speech, reasons for self-restraint must tackle hatred head on, while for unintended incitement to hatred such reasons can come from showing that the relevant harmful effects are sufficient to outweigh the reasons to speak. He then identifies and analyses three 'codes' that reasons for self-restraint can arise from: codes of civility, ethics, and morality. Each of these codes functions in a somewhat different way and gives different kinds of reasons for self-restraint, which might apply in different circumstances. Finally, Thompson turns back to the state by suggesting that this analysis could inform non-coercive state strategies aimed at combatting hate speech by encouraging and facilitating self-restraint.

In summary, each article in the special issue offers a distinctive contribution to the existing literature. Barendt and Whitten offer valuable critical insights into the post-Millian constitutive conception of harm that is central to the speech-act theory approaches that have become very influential in the recent literature. Richardson-Self and Gelber, though writing from different perspectives, both identify new types of wrongful speech, respectively cis-hetero-misogynistic speech and anti-Western terrorist-extremist speech. In this way they contribute to complexifying existing analyses, suggesting that the identification of more fine-grained categories of speech might be a valuable endeavour in future research. Furthermore, by combining elements of earlier discussions of pornography and free speech with a sophisticated analysis of misogynistic hate speech, Richardson-Self's account reveals the continuity and profound links between those literatures.

While focusing on a well-known category of wrongful speech, i.e. religiously offensive speech, Bonotti and Seglow offer an innovative analysis by drawing on theories of self-respect and non-domination, which have not previously been applied to this area of inquiry. Reid's article is distinctive for two reasons: it focuses especially on the speech of politicians, and it does so by distinguishing two kinds of wrongful speech (hate speech and inaccurate speech) that are often not sufficiently disentangled. In the era of fake news, Reid's account offers a useful framework for critically assessing false claims in public debate. Finally, despite the many differences between them, both Billingham's and Thompson's papers focus on responses to wrongful speech that do not involve coercive state censorship. There is urgent need for more creative thinking on interventions to mitigate hate and other kinds of wrongful speech, and Billingham's and Thompson's contributions offer valuable insights in this direction.

Acknowledgements The idea for this special issue originated from a workshop on 'Religion, Hate and Offence in a Changing world' held at Cardiff University on 14-15 December 2016. Earlier versions of some of the papers in the special issue were presented at that workshop. The guest editors would like to thank the editors of *Ethical Theory and Moral Practice*, Marcus Düwell and Thomas Schramme, for their support, as well as all the anonymous reviewers who provided invaluable feedback on the papers.

References

- Baker CE (1989) Human liberty and freedom of speech. Oxford University Press, New York
- Barendt E (2005) Freedom of speech, 2nd edn. Oxford University Press, Oxford
- Brettschneider C (2012) When the state speaks, what should it say? How democracies can protect expression and promote equality. Princeton University Press, Princeton
- Brink D (2008) Mill's liberal principles and freedom of expression. In: ten CL (ed) Mill's on liberty: a critical guide. Cambridge University Press, London, pp 40–61
- Brown A (2008) The racial and religious hatred act 2006: a Millian response. *Crit Rev Int Soc Pol Phil* 11:1–24
- Brown A (2015) Hate speech law: a philosophical examination. Routledge, New York and Abingdon
- Dworkin R (1981) Is there a right to pornography? *Oxf J Leg Stud* 1:177–212
- Dworkin R (2009) Foreword. In: Hare I, Weinstein J (eds) Extreme speech and democracy. Oxford University Press, Oxford, pp v–ix
- Dyzenhaus D (1992) John Stuart Mill and the harm of pornography. *Ethics* 102:534–551
- Feinberg J (1985) Offense to others (the moral limits of the criminal law Vol II). Oxford University Press, Oxford
- Hare I (2009) Blasphemy and incitement to religious hatred: free speech dogma and doctrine. In: Hare I, Weinstein J (eds) Extreme speech and democracy. Oxford University Press, Oxford, pp 289–310
- Heinze E (2016) Hate speech and democratic citizenship. Oxford University Press, Oxford
- Jones P (1990a) Rushdie, race and religion. *Political Studies* 38:687–694
- Jones P (1990b) Respecting beliefs and rebuking Rushdie. *Br J Polit Sci* 20:415–437
- Lægaard S (2007) The cartoon controversy: offense, identity, oppression? *Political Studies* 55:481–498
- Langton R (1990) Whose right? Ronald Dworkin, women, and pornographers. *Philos Public Aff* 19:311–359
- Langton R (1993) Speech acts and unspeakable acts. *Philos Public Aff* 22:293–330
- MacKinnon CA (1993) Only words. Harvard University Press, Cambridge, Mass
- Meiklejohn A (1961) The first amendment is an absolute. *The Supreme Court Review* 1961:245–266
- Mill JS (2006) [1859] On liberty and the subjection of women (ed. a Ryan). Penguin Books, London
- Parekh B (2006) Rethinking multiculturalism, 2nd edn. Palgrave Macmillan, Basingstoke
- Post R (2007) Religion and freedom of speech: portraits of Muhammad. *Constellations* 14:72–90
- Rostbøll CF (2009) Autonomy, respect, and arrogance in the Danish cartoon controversy. *Political Theory* 37: 623–648
- Scanlon TM (1972) A theory of freedom of expression. *Philos Public Aff* 1:204–226
- Scanlon TM (1979) Freedom of expression and categories of expression. *University of Pittsburgh Law Review* 40:519–550
- Sunstein CR (1993) Democracy and the problem of free speech. The Free Press, New York
- van Mill D (2018) Freedom of speech. In: Zalta EN (ed) The Stanford encyclopedia of philosophy (summer 2018 edition). Available at: <https://plato.stanford.edu/archives/sum2018/entries/freedom-speech/> (Accessed on 3 August 2019)
- Waldron J (2012) The harm in hate speech. Harvard University Press, Cambridge; London
- Warburton N (2009) Free speech: a very short introduction. Oxford University Press, Oxford

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.