16.

Hate Speech and the Distribution of the Costs and Benefits of Freedom of Speech

David Brax

Abstract
This chapter presents a broadly consequentialist argument in favour of hate speech regulations. The argument proceeds in two steps: First by exploring the familiar point that hate speech does harm by undermining the speech of targeted groups. And second, by considering the distribution of costs and benefits for allowing hate speech, which is unlikely to be fair or equal. The harm caused by hate speech primarily befalls people that are already among the worst off in our society. According to the principle that costs and benefits that befall those worst off matter more, morally speaking - a view known as 'prioritarianism' or 'the priority view' - distribution matters. I argue that this approach offers the most plausible argument in favour of hate speech regulations. Hate speech should only be unregulated if the predicted benefits are likely to outweigh the predicted costs, when the distribution of costs and benefits is also taken into account. It should be unregulated only if the predicted benefits are likely to outweigh the predicted costs, and the distribution of costs and benefits is also taken into account.

Keywords: consequentialism, fair distribution, free speech, liberalism, deliberative democracy, speech regulation, hate speech

The aim of this chapter is to provide a rather straightforward argument in favour of hate speech regulations. The argument is a highly general one, and as such can be applied in contexts spanning from campus speech codes and media guidelines to criminal law and international conventions and agreements (Such as the European Convention of Human Rights, and the Council Framework Decision on Racism and Xenophobia). The argument is broadly consequentialist: It is based on a cost-benefit analysis of restricted vs. unrestricted free speech. The argument can be roughly divided into two parts. The first is the familiar point that hate speech does harm by undermining the speech of targeted groups, which means that the utility that free speech exists to serve is diminished. The second part takes into account the *distribution* of these costs and benefits. The distribution of costs and benefits for allowing hate speech is unlikely to be fair or equal. The harm caused by hate speech primarily befalls people that are already among the worst off in our society. This, I argue, is a reason in favour of hate

speech regulations, even if there would be a net benefit in 'absolute' terms for allowing such speech. The argument thus depends on the intuition that equality has normative weight. It is, however, not based on the egalitarian principle that equality has *intrinsic* value. It is based on the principle that costs and benefits that befalls those worst off matters more. I argue that this normative theory, known as 'prioritarianism' or 'the priority view', offers the most plausible argument in favour of hate speech regulations.

Free speech and deliberative values

Respect for principles of free speech is a vital part of any well-functioning society. Arguably, it is even more vital in less well-functioning ones. That is when people, especially those in opposition and/or in minority positions, really need to be able to speak truth to power without fear of repercussions. In the ideal theory of a *deliberative democracy* associated with the works of John Rawls and Jürgen Habermas, free speech is key to citizen participation, which in turn functions to facilitate good political decisions, both in terms of being conducive towards the greater good and in terms of being legitimate in the eyes of those citizens. Free speech is intimately tied to the rights of individuals in relation to the state: people ought generally to be free to do what they want, as long as that behaviour does not harm others (Mill 1978 [1859]) or infringes on the equal freedom of others (Rawls 1971). The rights that a state needs to safeguard concern what people can and cannot do to each other.

While the more exact function and value of free speech are contested notions, it is at least in part dependent on the value of the activity in which it allows agents to participate. The value of freedom of speech is at least in part determined by the value of agents' abilities to engage in speech, and their ability to use speech to influence the conditions under which they live. In short, the value of freedom of speech is partly determined by the value of autonomy.

Most accounts of the value of freedom of speech recognise that freedom, which we may have a right to, occasionally comes into conflict with other values. In *On Liberty*, Mill describes the conflict between freedom and the authority of a state (Mill 1978, [1859]). Others point to the conflict between freedom of speech and equality (Fiss 1996, Brink 2001, Svensson & Edström, 2014), or with the protection of dignity (Waldron 2012). Utilitarians argue that the value of freedom depends on its utility. While freedom makes the effective pursuit of happiness possible, it does not secure it. Freedom in general is compatible with a great variety of outcomes. If the value of free speech is purely *instrumental* it can be evaluated accordingly. Whether or not hate speech regulation is warranted then depends on whether it can be shown to diminish the harm caused by hate speech while having no (or acceptable) detrimental effects on the utility of free speech in general.

In the minimal sense, freedom of speech is merely the absence of censorship, but there are more or less broad senses that are of greater interest (see Kenyon 2014, for

a discussion of the positive, as opposed to negative notion of freedom of speech. See also Karppinen in this volume). Deliberative democracy relies on voices being heard, yet, the right to vote aside, there is no *right* to be heard corresponding to the freedom of speech. The fact that you are allowed to say what you want does not correspond to a duty for others to take your opinion into account in their own deliberative processes. It merely *allows* for that to happen. In order for free speech to realise its full potential, then, the conditions for participation need to be favourable, and this may go way beyond the mere absence of censorship. Restrictions of free speech can therefore be justified with appeal to the values served by freedom of speech itself. Being free to say what you like does not guarantee that your speech gets a fair hearing, or that your influence in the deliberative process is determined by the quality of your argument. There is a large set of conditions and restraints, as reflected throughout this volume. Even if we were to accept the 'market place of ideas' metaphor for speech, implicit in the works of John Stuart Mill and explicit in a famous statement by Justice Oliver Wendell Holmes (1919), people do not start out on an equal footing, and the conditions can hardly be described as fair. A person, a politician, say, with a large budget will have a much easier time of getting his/her point across than a politician with more modest means. In the US, famous for its commitment to free speech, this fact has led to an intense debate regarding campaign financing (see Sunstein 1993). Inequality is built into this process, meaning that different agents have different capacities to begin with. The value of freedom is arguably based on the value of giving everyone an equal chance to succeed in his/her projects, but there is a risk that freedoms combined with inequality at the outset may lead to increased inequalities. Mere procedural equality need not serve the deliberative process and may even lead to discriminatory outcomes. In this regard, equality as an ideal may easily come into conflict with freedom as an ideal. If equality is of value, this is one reason in favour of restrictions and regulations. However, such measures are particularly controversial when it comes to speech. Whereas many restrictions on speech exists, in particular in advertisement and broadcast media, any restrictions on speech are always a matter of concern

Hate speech and harm

The aim of this chapter is to present a utilitarian argument in favour of hate speech regulations with appeal to the harmful consequences of hate speech. As mentioned, the argument is intended to be highly general, and thus not tailored to defend any specific item of hate speech regulation. For the purposes of this chapter 'hate speech' is understood quite broadly as speech that targets people based on group characteristics and portray the members of that group as "not worthy of equal citizenship" (Waldron, 2012). Whereas hate speech thus defined potentially cover all types of groups, it is particularly harmful when it targets disadvantaged groups, and we may therefore decide to narrow the scope of a hate speech regulation so that it protect only such

groups. There are issues concerning how to define 'disadvantage' on a group level, however, that I will sidestep for now. I take it for granted that words can do harm in a broad sense, and that this harm goes beyond the mere taking of offense. The way words hurt is familiar from studies of bullying, harassment, threats, provocation, libel, defamation and the "infliction of emotional distress" (Delgado and Stefancic 2004; Fiss 1996). Among the effects of being victimised by hate speech is the tendency to withdraw from social and public life, which means lost opportunities for interactions and social and economic loss. Of course, not every instance of what qualifies as hate speech has this effect. But, as Jeremy Waldron points out: hate speech can be understood along the lines of pollution, or like a slow-working poison (Waldron 2012:96). Hate speech is, in effect, polluting the social environment. More specifically, the harm of hate speech which is of particular interest here, considering the argument in the previous section, is the detrimental effect on the speech of others: its effect is to (and often intended to) silence them.

The argument is broadly Millian in nature: it connects the value of free speech to deliberative values, but in a purely instrumental manner. Deliberation serves autonomy, which in turn serves the successful pursuit of happiness. Mill is a utilitarian, after all (see Brink 2001; Sunstein 1993). Freedom of speech is supposed to secure the availability of diverse views and arguments from which citizens are able to make informed decisions. If we thus treat the value of freedom of speech as instrumental, and some modes of speech can be shown to do harm by undermining this function, we can make an argument in favour of restricting it, and claim that such speech is not worthy of protection. If speech is merely 'formally' free, and fear and disadvantage constrains what voices are being heard, rectifying this state of affairs may very well be in the state's legitimate interest. Hate speech in the sense regulated against in most European countries contributes very little to the furtherance of the ends of deliberation and is largely detrimental to it – and to the extent that it does contribute, it *is* protected by most hate speech regulations (see Bleich 2011).

On legal and moral wrongs, and the utility of freedom

While it seems quite obviously morally wrong to engage in hate speech, we have the right to do some things that it is clearly wrong for us to do. Ideally, we would not need to restrict free speech, because people would refrain from harmful speech on their own accord. But the freedom to behave badly may be an important freedom: to recognise this is what valuing autonomy is all about. The legitimacy of speech regulations hinges on how narrowly we can tailor these laws to target harmful speech without having a detrimental 'chilling' effect on open debate. But it also depends on the value we assign to letting moral behaviour develop as informed by non-legal reasons. It should be recognised, however, that such considerations have not stopped us in general from taking up legal measures when social norms fail to keep people from harming

each other. Taking the harms of hate speech seriously at the very least requires taking regulation into consideration despite its clash with freedom.

In a recent paper Marcus Schulzke (2015) offers a version of a typical consequentialist argument for protecting hate speech. He does so by appeal to the social benefits of exposing prejudices. Protecting hate speech may facilitate societal trust by allowing a broader range of views to be expressed, and it also gives an opportunity for 'counter speech' that in turn can influence the views of the speaker. This argument, then, depends on the estimate that hateful views will continue to exist and do harm even if they are no longer expressed due to fear of punishment. It also depends on the estimate that hateful views will tend to be successfully countered. This is an interesting argument. It is, however, difficult to believe that people harbouring such views would remain quiet if faced with hate speech bans. As noted above, most hate speech laws are quite narrowly tailored to target those modes of speech that are likely to do harm. The question, which is admittedly open, then is if allowing hate speech would have the positive effect that Schulzke projects, and, if it does, if it would be strong enough to outweigh the loss of speech resulting from being victimised by hate speech, mentioned above. I will return to this in the next section, which deals with the consequentialist calculus.

While the best response to hate speech may be 'more speech' and 'counter speech', this solution is not always available, is not always available to everyone, and is not always effective. Indeed, it is precisely because there are people who will have nobody standing up for them that there is a need for law in this and similar matters. Just as Schulzke argues, it is important that these views are met and argued against, but in fact, there is no evidence of decline in the discussion about racism and bigotry in countries that carry hate speech laws. It is primarily under circumstances where the effect of hate speech is to silence targeted individuals and groups that regulation may be called for.

To some extent, the critics of hate speech regulations are right: These regulations are intended to have a 'chilling' effect on public speech. The intention is to chill speech that is harmful in a sense similar to that of libel or defamation. There are direct and indirect, short- and long-term effects on the targeted communities and the social standing of the people belonging to those communities. The intention behind hate speech regulations is, among other things, to have a chilling effect on speech that has a chilling effect on *other* (high-value) speech. These regulations, then, are motivated by the same deliberative values that favour the protection of free speech in the first place. They are aimed at securing, rather than limiting, the availability of a broad spectrum of ideas and opinions. It should be noted that this mode of reasoning is strictly viewpoint neutral (see Sunstein 1993). Hate filled content is often the *medium* through which these harmful consequences are brought about, but the reasons for regulation are neutral insofar that *any* content with the same sort of effect could be banned on these grounds. In jurisdictions that carry hate speech provisions, such as those existing in most European countries, there is normally an exception made for truly deliberative contexts (see Bleich 2011). This means that hate speech laws are rarely blanket bans on expressing certain view-points; in a context in which racism,

for instance, would take the form of an articulated point of view, such speech would not be banned.

A second argument against hate speech regulations is based on the fear of government intervention, and the suspicion that allowing speech restrictions sets a dangerous precedent. While current governments may introduce and courts uphold such laws with good intentions and to good effect, there is always a risk that subsequent governments will abuse such laws to silence critics. This broadly libertarian argument has some merit, and it is clearly in the public interest to carefully tailor such laws in a manner that will leave minimal room for abuse. However, this risk must be weighed against both the abuse made possible by the *lack* of regulation *and* the likelihood that an oppressive government will find ways to infringe on the liberties of its citizens even in the absence of hate speech laws.

The relevance of the distribution of costs and benefits

This last section is devoted to the more precise normative basis of hate speech regulations. Freedom is often posited against other values such as equality, utility and the freedom of others. As mentioned, freedom may *serve* other values by making it possible for people to pursue them. But it does not *secure* such values. A free market can serve the wealth of a nation, but it offers no *guarantee* that the wealth is maximised, and certainly no guarantee that the wealth is distributed equally or fairly. Freedom can typically be limited when there are other compelling interests at stake. Regulations of markets arguably exist in order to heighten the probability that utility is maximised and/or distributed in some gainful manner.

Let us now assume that the benefits of a relatively unregulated freedom of speech are considerable, and that the costs considered above do not outweigh the benefits. This argument relies on the claim that any attempt to curtail speech, even if restricted to the kind that typically has harmful effects, will have a negative net worth. Arguably, if hate speech could be somehow isolated, its net contribution would be on the cost side of the calculus. The argument against legislation, then, must be that the benefits of free speech in general would be undermined if an exception was made and hate speech was regulated. For the sake of the argument, let us assume that this holds: the net worth of free speech where there are no hate speech restrictions is greater than the alternative.

What about the fact that the cost and benefits are not distributed *equally*? There are two considerations that would then apply: One is that equality (or fairness, if you prefer) might be of value in itself. The other, which is the argument that I put forward, is that distribution matters in the utility function. According to the view called 'prioritarianism', or 'the priority view' (Parfit 1997) more weight should be given to costs and benefits that befalls those that are worst off in a society. So even if there is a net benefit of unrestricted free speech, we must take into account whether the worst off

benefit or whether they are stuck with most of the costs. And if they do suffer most of the costs, this should be given more weight than the benefit for those that are better off at the outset. The advantage of this theory over egalitarianism is that it avoids the 'levelling down'-objection: that is, it does not say that you can improve on a situation merely be bringing those best off down a peg, which seems to be a consequence of intrinsically valuing equality.

The result of a prioritarian approach is that, other things being equal, a cost should be given greater weight if it disproportionally affects those worst off in a society. This effect, mind, can take place even if the immediate target of hate speech is not him/herself in a particularly disadvantaged position – indeed, there is a tendency for hate speech to target those individuals who, despite belonging to vulnerable and normally silenced groups, have achieved some sort of societal status. Whereas the harm in such cases does not befall a person that is among the worst off, the harm befalls a group that is. These groups, because of their marginalised position, are at a general societal *disadvantage*, and thus not in an ideal position to engage in successful counter-speech (Wolff and De-Shalit 2007). An argument in favour of unrestricted speech would, on this theory, need to be one that was acceptable to those that carry a disproportionate part of the costs. But if hate speech has such a detrimental net effect on those worst off, there is a strong case in favour of legislating against it, even if you believe that the benefits of unrestricted speech outweigh the costs in absolute terms. Note, however, that this is still a matter of consequentialist calculus. It does not mean minority interests will always trump majority interests. It merely means that in this calculus, the interests of those worst off count for more.

Concluding remarks

We have a right to behave in ways that it may be wrong for us to do. Morally, I should be chided for disrespecting people, but I should probably not be legally prohibited from doing so. While morality and law are intimately related, they do not coincide. Having a sphere of optional actions, even actions with moral importance, is crucial for human flourishing in general. The limits concerning what we should be allowed to do to each other are arguably given by the rights of others, primarily the right not to be harmed. People often point out that we do not have a right not to be offended, and that hate speech laws wrongly imply that we do. The same people often recognise that we have a right not to be threatened, libelled, bullied, perhaps even a right not to be humiliated. Given that the accumulative effects of hate speech have effects similar to threats, libel and bullying, the case for legislating against it should be given a fair hearing.

The argument put forward in this chapter is that the harms of hate speech should be given particular normative weight when it hits those that are worst off in society. This is a broadly consequentialist view insofar as the value of free speech is given by a utility function. It should be unregulated only if the predicted benefits are likely to

outweigh the predicted costs. But the weight of those consequences should take the distribution into account.

References

Bleich, Erich (2011). *The Freedom to be racist – How the United States and Europe Struggle to Preserve Freedom and Combat Racism*. Oxford: Oxford University Press.

Brink, David O. (2001). 'Millian principles, freedom of expression, and hate speech', *Legal Theory* 7(2):119-157.

Council Framework Decision 2008/913/JHA of 28 November 2008 on combating certain forms and expressions of racism and xenophobia by means of criminal law.

Delgado, Richard & Stefanicic, Jean (2004). *Understanding Words that Wound,* Boulder CO: Westview Press.

Fiss, Owen M. (1996). *The Irony of Free Speech.* Cambridge, MA: Harvard University Press.

Hellman, Deborah (2011). *When is discrimination wrong?* Cambridge, MA: Harvard University Press.

Holmes, Oliver Wendel (1919). *Abrams v. United States*, 250 U.S. 616, 630.

Kenyon, Andrew T. (2014). 'Assuming Free Speech', *The Modern Law Review*, 77(3):379-408.

Lewis, Anthony (2009). *Freedom for the thought that we hate: A Biography of the First Amendment*. New York: Basic Books.

Mill, John Stuart (1978, [1859]). *On Liberty*, Indianapolis: Hackett Publishing

Parfit, Derek (1997). 'Equality and Priority', *Ratio* 10:202-221.

Schultze, Markus (2015). 'The Social benefits of Protecting Hate Speech and Exposing Sources of Prejudice', *Res Publica* online Oct 2015:1-18. DOI 10.1007/s11158-015-9282-1.

Svensson, Eva-Maria & Edström, Maria (2014). 'Freedom of Expression vs. Gender Equality', *Tidskrift for Rettsvitenskap* 127:479-511.

Sunstein, Cass (1995). *Democracy and the Problem of Free Speech*. New York: Free Press.

Rawls, John (1971). *A Theory of Justice*. Cambridge, MA: Belknap Press of Harvard University Press.

Waldron, Jeremy (2012). *The Harm in Hate Speech*. Cambridge, MA: Harvard University Press.

Wolff, Jonathan & De-shalitt, Avner (2007). *Disadvantage.* New York: Oxford University Press.