

'Speech which spreads hate should never be tolerated.' Discuss.

Although government ought to place some legal restrictions on hate speech because of the harms it causes, not all such hateful speech should be banned. Toleration of propositionally expressive hate speech, at both the legal and social level, is important for the ability of citizens to participate in democratic discussion, and ultimately for social welfare. In this essay, I first provide a working characterisation of what it means for speech to spread hate, and clarify what is meant by "toleration". I then present the consequentialist case for restrictions on hateful speech, identifying the empirical premisses which must be met for the argument to hold. I then demonstrate that some restrictions on hateful speech should be endorsed even by those who believe that free speech is intrinsically valuable, or that it is a means to ends other than those of utility (for example, autonomy). Finally, I conclude that whilst the fact that some particular speech spreads hate is a *pro tanto* reason for its prohibition, this interest in direct harm prevention may sometimes be outweighed by the indirect harm done by intolerance of others' speech.

Defining hate speech is an important first step in any discussion of what our policy towards it should be. As Anderson and Barnes (2022) note, whilst hate speech could be identified in terms of the effects it has on the targets, this results in a definition which is vague and under-specified. Moreover, using this as the basis to defend prohibitions on hate speech seems to put the cart before the horse, in that it assumes there unquestionably are great harms being done by this speech. Therefore, I follow the approach taken by Parekh (2012, 40), centred around the content of the speech in question: hate speech is that which "expresses, encourages, stirs up, or incites hatred against a group of individuals... based on an arbitrary and normatively irrelevant feature". Merely articulating one's mistrust or dislike of such a group would not qualify as hate speech.

Toleration, in this context, refers to the absence of legal interdictions and social penalties against certain speech. These forms of prohibition and intolerance are closely related, and it would be a mistake to draw a false dichotomy between the two: whilst the magnitude of costs facing an individual who decides to break the law may be significantly larger than those borne by one who decides to break with social convention, the qualitative effect is the same (Mill 1859, 63). Social tolerance does not require that we sit mute whilst others spout what we believe to be bigoted falsehoods, but it does entail that we refrain from launching *ad hominem* attacks or seeking to make their life more difficult as a result of airing those views. (Ibid, 144). Of course, identifying whether tolerance is permissible (or required) in the case of hate speech is the goal of this essay.

For those with a utilitarian moral framework, free speech and tolerance are both only valuable insofar as they promote the general welfare. We can therefore construct a consequentialist case for prohibitions on hate speech as follows:

- P1:** Hate speech does harm to those it is directed at.
- P2:** Ceasing to tolerate hate speech would reduce its incidence.
- P3:** Any costs to aggregate welfare from ceasing to tolerate hate speech are outweighed by benefits from reducing the incidence of such speech.
- C:** Society should cease to tolerate hate speech.

Few would reject **P1**. As well as the direct psychological harms inflicted on the targets of hate, such speech also has pernicious longer-term effects, creating conditions in which groups of society are more likely to be ostracised, discriminated against, and subjected to violence (Parekh 2012, 44-45). **P2** and **P3** are more contentious, though ultimately empirical, premisses. Opponents of restrictions on hate speech marshal a variety of arguments against **P2**, though perhaps the two most common objections are (a) that intolerance of hate speech in the public sphere would only cause it to instead fester underground,

Commented [1]:
Good

Commented [2]:
Another strong introduction. I suppose this will follow, but I wonder here exactly what the indirect harm is in these cases. Is the idea that the intolerance of some expressive acts is itself harmful in some way? Or that the interest in preserving free expression in some cases outweighs (for whatever reason) the harm that it causes? If the latter, then this could be made a little clearer.

Commented [3]:
It's been a while since I read this paper but I've always wondered what, exactly, makes a feature normatively irrelevant in this way? Pointing to its arbitrariness usually invokes the idea that it is not a justifiable basis for discrimination but this also seems to imply, in some way, that there are non-arbitrary and normatively relevant features. Is it the fact that some hate is arbitrary in this way that makes it so harmful or simply the fact that it is hateful? Or some combination of these things?

Commented [4]:
Could there be a context in which it would?

Commented [5]:
Good - although does the inclusion of social penalties broaden things somewhat? I think it seems sensible to include this where punitive social consequences are so high such that there is a de facto prohibition on the speech but I suppose it depends on the kind of penalty and the state's willingness to intervene to protect free speech whenever the social costs are particularly high so as to act as a form of coercion. To a large extent, this also comes down to what we think freedom consists in.

Commented [6]:
Is this too inclusive? Some people characterise tolerance as benign neglect so I suppose this kind of response might fall foul of that.

Commented [7]:
I'm wondering if this is better phrased as a conditional argument "If any costs to aggregate welfare are outweighed.... Then ... society should". Perhaps you are setting this up just to knock it down but most utilitarians would (I think) endorse the conditional form. Though I suppose this is what you are pointing to in the introduction when discussing the empirical reality of whether certain conditions are met.

Commented [8]:
Doesn't this depend on what we say "hate" speech is though? And even when we agree on what is hateful, is it the case that everyone is harmed by it (even those it targets)? Some people will also be harmed to different degrees etc.

and (b) that the most effective way to root out hate is through debate and discussion (Baker 2009, 155). The case for (a) looks especially weak: if hate speech alone occurs underground, then it inflicts no harm on targets; equally, if this speech leads to discrimination and violence, then the government should additionally act to limit and punish this further law-breaking (Anderson & Barnes 2022). As for (b), the following series of events illustrates one instance where further debate and discussion cannot be an effective way of reducing the incidence of hate:

Stephen launches a hateful series of attacks against Ruby, and everyone else with red skin. Ruby experiences psychological harm, and her attempts to calmly rebut the vilification simply lead Stephen to engage in more diatribes. As a result, Ruby and other reds withdraw from the public sphere, in the hope that Stephen is less likely to attack them.

Just as government intervention in some sections of the economy is necessary to prevent market failure and collapse, so too is regulation of the marketplace of ideas. We can formalise this by appealing to the concept of speech acts. In the context of pornography, the speech of pornographers limits women's perlocutionary and illocutionary power, by making it more likely that men ignore them when they say "no", and by circulating the notion that "no" really means "yes", respectively (Langton 1993). More generally, hate speech inhibits others' ability to enter, and be heard in, the public square. This is directly analogous to how, in our definition of tolerance, we recognised that overbearing social penalties reduce individuals' ability for free speech.

Once the issue of hate speech is viewed from this angle, the outline of a rights-based justification for intolerance of hate emerges. For those who believe that free speech is not simply a means to welfare, the consequentialist argument will not hold – it is not enough to present on a balance-sheet the costs and benefits if free speech is viewed as something of special worth (Dworkin 2009, vii). Yet Langton's argument cuts through this, by showing that hate speech poses a threat to speech more widely, and not just to the distinct goods of equality and dignity. Indeed, even if one believed that liberty of expression had lexical priority over any other intrinsic good, one would have to concede that there is at least a potential case for restrictions on speech that undermines others' free speech.¹

Returning to the consequentialist argument for hate speech restrictions, what about the soundness of P3? A free speech absolutist may attempt to undermine the claim that intolerance of hate is beneficial overall by pointing to the risk of stumbling down a slippery slope of censorship. Sufficiently narrow framing of the scope of laws against hate speech might mitigate against this risk: we already have bans on libel and defamation, but not against criticism or petty insults (Anderson & Barnes 2022). However, that seems to suggest that perhaps some hate ought to be tolerated, as part of the price for avoiding an undesirable chilling effect on legitimate speech. Although it certainly seems naïve to think that more speech, no matter how hateful, is *always* ultimately better, that is a far lower bar than a proponent of the titular statement must clear – their task is to show that, if it includes any content spreading hate, tolerating more speech is *never* better. Let us now examine more closely the plausibility of that claim.

There are some kinds of speech which spread hate by propagating negative claims about groups of individuals, but do not explicitly vilify the targets – for instance, statements like "All reds are dirty and uneducated criminals!". Against the view that speech which spreads hate should never be tolerated, I suggest that we should, in fact, tolerate speech of this nature, because it inflicts only limited harms on the targets, and does not adversely affect their ability to exercise freedom of speech. The distinguishing feature of this permissible speech is that its content is primarily propositional. That is, it communicates objects of belief which are either true or false. Whilst the targets of such speech will no doubt still be

¹ I am careful to state there is only a "potential" case. It could be true that restrictions would infringe upon the liberty interests of the speaker more than they would protect the liberty interests of the target – but this would an empirical matter to be resolved with a specific problem at hand.

Commented [9]:
I am not very sympathetic to Baker but I take it that his point is more that the prohibition of speech may actually lead to the proliferation of hate amongst certain groups and now that we've outlawed the speech, it is less easy for us to monitor and engage with counter-speech.

Commented [10]:
Good

Commented [11]:
Good - but I wonder whether the consequentialist argument could be steelmanned to reach a similar conclusion (especially in the conditional form)? It just seems like some expressive acts/debates will be of very little utility even as part of the 'marketplace of ideas' whilst also causing a lot of psychic or even physical harm.

harm, the extent of this is greatly reduced by the fact that they are able to challenge and refute the propositions in question with rational argument. As Yong (2011) puts it, “[b]ecause assertions of fact and evaluative opinions have clear cognitive content, they can appropriately be answered through deliberative and articulate speech”.

This holds whether one adopts a consequentialist or rights-based position in assessing the value of free speech. From a rights-based perspective, it seems unlikely that this speech significantly impairs the targets’ right to exercise their own speech, or that their interest in being treated with dignity as equal members of society is adversely affected, when considered relative to the claim that speakers have to speak. Equally, taking a consequentialist standpoint, the damage done by undermining the principle of free debate would be greater than that which would be avoided by banning this form of speech. Again, there is a clear parallel with our understanding of tolerance, grounded in Mill’s view of liberty: people must be able to robustly criticise their intellectual opponents, but may not use rhetoric to inflict harm on someone because they are an intellectual opponent; similarly, people must be able to robustly criticise other groups of citizens, but may not use rhetoric to inflict harm on someone because they are part of a certain group of citizens. Identifying the motivations behind an individual’s speech is impractical, but tolerating only hate-spreading speech whose content is primarily propositional is an eminently workable operationalisation of this principle.

So, to conclude, it is not true that speech which spreads hate should never be tolerated. The fact that some speech expresses or causes hate provides a *pro tanto* – but not indefeasible – reason for it to be regulated. This is because although hate is harmful (through both its direct and indirect effects on victims, which include a reduction in their ability to exercise free speech), so is intolerance of hate speech. I have argued that, on consequentialist and rights-based justifications for free speech, there are strong reasons for hateful speech that is mostly non-propositional to be regulated, but for speech which indirectly spreads hate via the expression of propositional statements to be tolerated. Such an approach strikes the right balance between society’s interests in the prevention of harm and protection of liberty.

Bibliography

Anderson, L., & Barnes, M. (2023). Hate Speech. In E. N. Zalta & U. Nodelman (eds.), *The Stanford Encyclopedia of Philosophy* (Fall 2023 Edition). <https://plato.stanford.edu/archives/fall2023/entries/hate-speech>.

Baker, C. E. (2009). Autonomy and hate speech. In I. Hare & J. Weinstein (eds.), *Extreme Speech and Democracy* (pp. 139–157). <https://doi.org/10.1093/acprof:oso/9780199548781.003.0009>.

Dworkin, R. (2009). Foreword. In I. Hare & J. Weinstein (eds.), *Extreme Speech and Democracy* (pp. v–x). <https://doi.org/10.1093/acprof:oso/9780199548781.002.0004>.

Langton, R. (1993). Speech Acts and Unspeakable Acts. *Philosophy & Public Affairs*, 22(4), 293–330. <http://www.jstor.org/stable/2265469>.

Parekh, B. (2012). Is there a case for banning hate speech? In M. Herz and P. Molnar (eds.), *The Content and Context of Hate Speech* (pp. 37–56). <https://doi.org/10.1017/cbo9781139042871.006>.

Mill, J. S. (1859). *On Liberty* (ed. Himmelfarb, G., 1985). Penguin.

Yong, C. (2011). Does freedom of speech include hate speech? *Res Publica*, 17(4), 385–403. <https://doi.org/10.1007/s11158-011-9158-y>.

Commented [12]:

I wonder if the speech above does not explicitly vilify its targets. Imagining the same kind of statement made about different ethnic groups is certainly disconcerting irrespective of whether there is a proposition that can be refuted (also, what about if the proposition is true but still liable to incite harm?). I think more could also be said about the nature of the harm that is being limited. For an argument of this kind to pass, I imagine that you’d need to take an objective account of some kind as some people would find such statements very harmful. Finally, though the content here is clear, I’m less sure of its potential epistemic value.

Commented [13]:

Doesn’t this depend on the circumstances? e.g., imagine this spray painted as a slogan all over town. Would Reds feel safe going out in public, airing their views, etc? Would this campaign be likely to incite hate/harm? What if it was combined with a political campaign of some sort? If people do have a pro tanto right to express themselves, then doesn’t this in part depend on the value we think that the speech is promoting? What underlying value is promoted in this case and is it sufficient to outweigh whatever is of disvalue here?

Commented [14]:

I’m not so certain that the proposition above does not fall foul of this requirement and you seem to acknowledge this above when you say that it only inflicts limited harms rather than no harm at all. Something else to consider is whether these kinds of statements regularly feature in an “intellectual debate” of some kind. It seems that the bar for such debates is usually set higher anyway such that this kind of speech is a rarity and is generally considered obstructive to productive debates

Commented [15]:

This could be delineated further - e.g., Tory politicians are a certain group of citizens - can hateful speech be permissibly be directed to them as Tory MPs? Or would this be normatively relevant in some cases?

Commented [16]:

I may be misunderstanding this a little but doesn’t the former sentence conditionalise things to include motivations? I.e., “may not use rhetoric... because they are” - this “because” seems to make it the case that they shouldn’t express harmful views on a certain basis. But doesn’t this then make a step to identify a motivation of some sort, or at the very least, an explanatory reason as to why agent A expresses view X.

Commented [17]:

Overall this is another really strong essay. Grade: 79