



ourcommunity.com.au

The free speech conundrum: Combatting hate and anger in the wild west of the internet

Presentation by

[Dr Anjalee de Silva](#)

Lecturer, Melbourne School, WLIA Fellow,
expert in anti-discrimination and free speech.

[Communities in Control Conference](#)

Moonee Ponds, Melbourne, and online,
Monday, May 29, 2023

Video and audio versions of this speech as well as Dr de Silva's
presentation slides, will be posted online at
www.communitiesincontrol.com.au

About the presentation

One outspoken billionaire tries to control the narrative in the guise of "free speech", and suddenly, Twitter accounts once banned for abuse and hate speech (including that of former US president Donald Trump) have returned to the platform. You only have to spend a minute on Twitter to see that its new approach to content moderation has already led to some of the most vile, offensive language imaginable.

Hate speech, abuse and anger are everywhere online. What can we do to combat them? How can we support those who deal with them?

Introduction by Denis Moriarty, group managing director, Our Community

It's now my pleasure to introduce you to Dr Anjalee de Silva.

Anjalee is a Lecturer at Melbourne Law School. She is an expert in anti-discrimination, free speech and media law and theory, with a focus on harmful speech and its regulation, particularly in online contexts.

In particular, her work examines vilification or 'hate speech' directed at and about women, as well as the role of law in deterring, regulating, and mitigating the harms of such speech.

Outside academia, Anjalee is a local government councillor at Monash City Council, where she also serves on the Gender Equity Advisory Committee.

Today, Anjalee will be speaking on something that I'm sure most people in this room have been forced to deal with: hate speech and anger in the wild west of the internet.

Please make her feel welcome.

Dr Anjalee de Silva

Language warning: Please note Dr de Silva's talk includes examples of abusive and offensive language.

Thank you so much, Denis, for that introduction, and thank you to our community for having me, and to all of you for being here as well. It's such a privilege and an honour to be here talking about this topic.

So, my topic today is the free speech conundrum, combating hate and anger in the wild west of the Internet. In speaking to this topic, my aim is to provide an overview of some key issues relating to online hate speech against women. Unfortunately, we don't have enough time to discuss all of the relevant details, but I do hope to provide an introduction that lends itself to a lively discussion afterwards if we do have time for questions.

My focus on women isn't only because women regularly bear the brunt of online hate and anger, both in terms of its volume and its severity. And it's not just because hate speech against women is my area of research expertise.

2

If quoting from this speech, please acknowledge that it was presented to the 2023 Communities in Control Conference convened by Our Community | www.communitiesincontrol.com.au



Communities in Control
The Conference. The Movement.



ourcommunity.com.au
Where not-for-profits go for help

Linking back to the conference theme this year, if a country does have a soul, surely, a large part of how that soul is constituted depends on the way in which we treat our women and girls. And it seems to me this problem of hate speech against women speaks both persuasively and poignantly to some of the things that might be ailing our soul as a nation, and indeed as a planet and as a race.

Having said this, online hate and anger, like most aspects of systemic discrimination, manifest and are experienced in a range of ways, and it's important to keep this in mind. Significantly, women themselves experience online hate and anger intersectionally along additional axes of systemic oppression that include race, colour, religion, sexuality, gender identity, disability, and so on.

Hate speech or vilification against women may be understood as discriminatory treatment that constitutes and causes the systemic subordination and silencing of women and girls on the basis of their actual or perceived female sex. This means that such speech is about all women, even when it's directed at particular women.

I use the terms 'hate speech' and 'vilification' interchangeably. I also use the terms 'sex' and 'sex-based vilification in favour of gender and gendered vilification' for a number of reasons. It's unclear whether gender expression is distinct from actual or perceived female sex or gender identity is an axis of women's systemic oppression in patriarchal societies in relevant ways for discussions around hate speech. For example, the vilification of women for their gender expression, including gender non-conformity, is an aspect of their vilification in patriarchal societies on the basis of their actual or perceived sex.

Sex-based vilification is also distinct from vilification on the basis of gender identity. Gender identity as a category of hate speech is typically addressed to hate speech directed at and about trans and intersex people for being trans and intersex. It excludes vilification directed at and about women, including transwomen, on the basis of their actual or perceived female sex.

So, hate speech against women is prolific on digital and online platforms. And the proliferation of digital and online media means that the prevalence and the severity of sex-based vilification is more easily observable and documented



than before. Recent examples of speech that might be characterised as hate speech against women cover everything from women's experiences of being victim to offhand sexist remarks, to revenge pornography, to invective directed at female journalists and bloggers, and to speech in the manosphere, which is a collection of websites, blogs, and other online forums promoting toxic masculinity.

The vitriol experienced by women with public profiles on Facebook and Twitter, on digital news media platforms, and in other online contexts, as well as the lesser known experiences of women without public profiles, speak to the prevalence of the problem. And a lot of you would have encountered mainstream media reports of some of this as well. The problem of hate speech against women is especially apparent in the context of the cyber harassment of women.

So, Professor of Law Danielle Citron defines cyber harassment as involving the intentional infliction of substantial emotional distress accomplished by online speech that's persistent enough to amount to a course of conduct rather than an isolated incident. So, the cyber harassment of women typically involves sustained and tactical campaigns engaging multiple forms of communicative conduct, including threats and violent invective, sexualised invective, including rape threats, non-consensual pornography, other objectifying speech, and other contemptuous speech. And I'm happy to speak to those categories in a bit more detail during the Q&A if they are of interest.

Cyber mobs of more than one assailant often engage in campaigns of cyber harassment against women. Anonymity and invisibility of assailants online, as well as the cross-jurisdictional and multi-jurisdictional nature of cyber harassment, makes it difficult to measure the extent of any given mob. Significantly, if individuals take part in mob-based campaigns, existing criminal laws may not be sufficient to catch them as being culpable as part of that mob. For example, their behaviour may fall short of thresholds for joint criminal liability or accessory liability under existing laws.

In addition, phenomena such as the lowering of individuals' inhibitions online and altered dynamics around interaction and decision-making online means that the harms of this kind of speech is accommodated, authorised, and magnified

4

If quoting from this speech, please acknowledge that it was presented to the 2023 Communities in Control Conference convened by Our Community | www.communitiesincontrol.com.au



Communities in Control
The Conference. The Movement.



ourcommunity.com.au
Where not-for-profits go for help

online. Online sex-based vilification is often directed at women in positions of political leadership or with public profiles.

In Australia, female politicians across the political spectrum have spoken openly about their many and varied experiences of communicative conduct that may reasonably be described as hate speech. So, Ditch the Witch was famously said of Julia Gillard while she was former Labour PM. Mehreen Faruqi, who is a Greens Party Senator, has written candidly of the intersectional and especially vitriolic sex-based vilification that she is subjected to as a Muslim woman of colour. Sarah Hanson-Young, also a Greens Party Senator, recently brought a successful defamation claim against a former male politician with respect to speech that also constituted hate speech. So, there is a sense in which some of this speech overlaps categories of harm as well. It might be hate speech, and it might also be defamatory speech.

Women with public profiles may be particularly targeted when they speak openly about issues affecting women. So, Julia Banks, for example, is a former Liberal Party Member of Parliament has observed that the online harassment that she experienced worsened after she publicly discussed her own experiences of sexism within Parliament.

International examples are also fairly prolific. So, Anita Sarkeesian, a Canadian-American feminist blogger and gamer, was targeted after starting a crowdfunding campaign to create a series of short films examining sexist stereotypes in video games. Caroline Criado-Perez was similarly besieged for heading up a successful campaign to have Jane Austen's image replace Charles Darwin's on the British 10-pound note. When Criado-Perez spoke about the abuse, including during media interviews, the campaign of invective against her escalated. And a number of high-profile women in the UK who pledged their support for her also started receiving floods of abuse.

So, it's relatively clear to see that attacks of this nature targeted at women politicians and other women advocating for women's interests have direct and clear impacts on our democracy. However, so too do attacks on ordinary women who don't have public profiles. The silencing harms of hate speech against women are particularly relevant in this regard.



Women typically feel threatened and humiliated by occurrences of hate speech, and they adapt their own behaviours accordingly. They police their identities, their speech, and their movements, or they leave online and offline spaces and disengage entirely from public life.

So, this kind of hate speech silences women by preventing them from speaking, by marginalising and devaluing their speech when they do speak, and in building structural constraints around their speech. The result is that, as I mentioned, even where women can and do speak, what they say is often unable to have its intended force. That is, hate speech against women functions and is often intended to exclude women from full democratic participation. And many of you may have encountered this occurring in your work, particularly if you pursue issues that are relevant to women. This is especially true of hate speech against women that occurs in spaces which are essential for political communication.

So, for many women, as for many others, online spaces are now key spaces of public discourse and engagement with public life. In liberal democracies, this is particularly the case. As such, women's presence in and engagement within those spaces pertains to democracy itself. It's important that women are able to be in these spaces and to engage as themselves.

If democratic legitimacy rests on political equality, meaning substantive equality in participating in the democratic process, hate speech against women delegitimises democracy and represents a crisis of democracy in itself. And this means that such speech warrants careful and urgent consideration in democratic societies and necessitates regulatory responses that appropriately and adequately address its harms.

Despite the prevalence and harms of this kind of hate speech, especially online, there is what could be called a sex-based gap in anti-hate speech laws. So, apart from some notable exceptions at the foreign domestic level, particularly South Africa and France, there is no jurisdiction currently that bans or prohibits in any way or mitigates in any way hate speech against women on the basis of their actual or perceived female sex.



In addition, this issue around hate speech against women has not received much scholarly or policy attention. In contrast, hate speech on the basis of other characteristics, including race, religion, sexuality, gender identity, intersex status, disability, and even HIV/AIDS status, is unlawful under international law and in many domestic jurisdictions. Considering these gaps as well as the prevalence of and difficulties in regulating hate speech against women online, a multifaceted approach is required to meaningfully address the harms of such speech.

So as part of such an approach, states would employ a range of legal strategies to respond to different kinds of online and offline sex-based hate speech, and law would only be one aspect of a holistic response that also incorporates other regulatory and non-regulatory counter-speech measures. So, counter-speech measures being measures that allow women to speak back on their own behalf and others to speak back on the behalf of women.

Speech acts constituting sex-based vilification may be regulated through a combination of content moderation laws and guidelines constituting a platformed response. So, content moderation laws and guidelines may be administered by state bodies, for example, through content moderation schemes, codes of conduct for social media firms, and other platform hosts. Australia's e-safety commissioner, for example, is legislatively empowered to negotiate directly with platforms for the removal of some material, including some material constituting hate speech.

Corporations and organisations - so for example, media and tech firms, including social media companies, Internet service providers and other platform hosts - may also be encouraged by states and by other actors to commit to voluntary codes of conduct or put in place internal guidelines for classifying, identifying, and removing this kind of speech. And then academics, lawyers, policymakers, including many of the people in this room, and others, may also work with platforms in various capacities to better the design of policies, procedures, and governance infrastructures relating to moderation of speech that is hate speech against women.

And additionally, counter-speech in all its forms is a really important aspect of any holistic response. So, in particular, platforms' non-regulatory contributions

7

If quoting from this speech, please acknowledge that it was presented to the 2023 Communities in Control Conference convened by Our Community | www.communitiesincontrol.com.au



Communities in Control
The Conference. The Movement.



ourcommunity.com.au
Where not-for-profits go for help

to educational and capabilities-building resources enabling women to themselves speak back or to enable others to speak back on behalf of women are really important measures as part of any holistic approach.

Finally, any platform measures taken as part of such an approach would need to be interpreted and applied by content moderators, and in some case, automated decision-making systems as algorithms increasingly are used to pick up and regulate this kind of speech. So, any such measures should for these reasons be accompanied by holistic and effective enculturation processes directed at their proper interpretation and their application.

And an example of this is Facebook's hate speech community standard, and a warning that this part of my presentation comes with a language warning. I can't get around it - I'm sorry - when I'm talking about this topic. So, Facebook's hate speech community standard primarily defines hate speech expressively rather than functionally. It defines it according to particular words that aren't allowed or particular phrases that aren't allowed as opposed to the kinds of things that hate speech ends up doing for those who are targeted by it. This is too prescriptive a way to go about it.

So, for example, the community standard says that it's prohibited on Facebook as a platform to refer to women as objects or household objects or property or compare them to these items in that way. But we all know that this is too narrow a standard, and that's not really how hate speech against women works. It's more dynamic and it's much slipperier than that in terms of both the language and the context that it employs to demean women.

So similarly, Facebook prohibits explicit terms such as 'whore' and 'slut,' but it prohibits the use of those terms to describe both men and women. Now, while it might be unpleasant to hear both men and women being described in this way, the impact on men is entirely different to the impact on women. The hate speech standard that Facebook employs by terming it as sex-neutral or as gender-neutral overlooks that male sexuality is not constructed as a source of shame for men as female sexuality is for women, and that male sexuality is rarely, if ever, commented on in comparable terms in patriarchal societies.



So, in other words, the community standard broadly doesn't reflect that contemptuous speech directed at and about men on the basis of their male sex doesn't and cannot systemically harm them in the ways that sex-based vilification harms women in patriarchal societies. And it's important that platforms' content moderation policies and processes engage with this level of complexity and nuance in consultation with experts.

It's also especially important with respect to sex-based vilifications that platforms don't, through policy oversight or overly narrow administration, reinforce the sex-based gap in law and policy relating to hate speech regulation that I described earlier. So, this danger was highlighted as part of a Facebook oversight board case recently made available for public comment. The case involved a decision by Facebook to remove a post containing a video in which the horrible term 'fag' was used.

However, the term 'bitch' was also used in the video as part of the phrase 'son of a bitch,' which is a very everyday phrase that we often hear. But the latter term and the associated phrase were not emphasised by the board as a subject of their decision as requiring a comment, even though it's a word and a phrase regularly deployed to demean women in a particular way.

So, in being able to identify and regulate the homophobic speech that occurred as part of that post but not the misogynistic speech that went hand in hand with it, the community standard and the way that it was enforced by the Facebook oversight board essentially reinforced this gap in hate speech law and regulation that deems hate speech against women not worthy of oversight and not worthy of sanction.

This highlights the broader issue, I think, that hate speech against women on the basis of their actual or perceived sex and also on the basis of their gender expression and gender identity in some contexts, particularly as regards the intersectional harassment that transwomen often faced, is generally speaking not only ubiquitous in the sense that it's everywhere all the time, but it's also invisible in the sense that it's extremely normalised. The treatment of women as inferior or as sexual objects, for example, may be so central to the way that we organise our societies in patriarchal societies that, unlike racist or homophobic



speech or other categories of hate speech, it's imperceptible as harm or it's imperceptible as harm worth doing anything about.

And this phenomenon, the simultaneous ubiquitousness and invisibility of this kind of hate speech, may partly explain the failure of platforms like Facebook to appropriately and adequately identify and respond to online hate speech against women, either in policy or in practice.

Again, it's important that this doesn't continue to happen, and platforms must work with the relevant experts, many of whom are in this room, to train their moderators as well as their algorithms to be sensitive to this kind of hate speech in the range of ways in which it manifests and which it harms women. As I flagged at the outset, women often bear the brunt of online hate, both in terms of volume and in terms of virulence, and it's imperative that such speech is addressed.

ENDS

MORE INFORMATION

For reports, audio, transcripts and video from the 2023 Communities in Control conference and from previous years, visit www.communitiesincontrol.com.au.

