



Monitoring Digital Hate: What the Christchurch Massacre Taught Us About the Limits of Free Speech

By Tracey Nicholls



KGS 2019

Dateline: CHARLOTTESVILLE (VA), USA, August 11, 2017 - A gathering of self-identified “alt-right” protestors marches through a park in this small college city waving white supremacist and Nazi-affiliated flags, chanting slogans identified with “white power” movements and so-called “Great Replacement” beliefs put forth by Islamophobes (“you will not replace us”) and slogans identified with Nazi ideology (“blood and soil”). In the name of (white) American history, they are protesting the planned removal of a statue of the general who led the army of the Confederate States of America, the Southern separatist movement that took up arms against the American government in the country’s 19th century Civil War (1861-1865). Subsequent protests result in beatings of counter-protestors and one death. Days later, the President of the United States, Donald Trump, notoriously defends the white supremacists by observing that there were “very fine people on both sides.” The organiser of this “Unite the Right” protest is known in Charlottesville for his sustained online harassment campaigns against city councilors who support the removal of racist monuments.

Dateline: CHRISTCHURCH, New Zealand, March 15, 2019 - An Australian man living in New Zealand attacks worshippers at two different mosques in the city of Christchurch, killing 51 and wounding many others. He is a proponent of the Islamophobic, anti-immigrant views of a global

“white power” network that disseminates its rhetoric of hate and its narrative of an imperiled white race online, via unregulated spaces within “the dark web” and via encrypted social media apps. His attack on Muslim New Zealanders is met with shock and grief within the country, an outpouring of solidarity that is expressed by Prime Minister Jacinda Ardern in her immediate response: “They were New Zealanders. They are us.” Of the shooter, whom she consistently refuses to identify by name, she says, “He is a terrorist. He is a criminal. He is an extremist...He may have sought notoriety, but we in New Zealand will give him nothing. Not even his name.” After the attacks, it becomes clear that he had been announcing his intentions in online forums and had been livestreaming the attack through a Facebook link. New Zealand moved swiftly to criminalise the viewing or sharing of the video of the attack.

Dateline: PARIS, France, May 15, 2019 - Two months after the Christchurch attack, New Zealand’s Prime Minister Jacinda Ardern stands at a lectern in a joint press conference with French President Emmanuel Macron to announce a non-binding agreement dubbed “The Christchurch Call to Action.” The agreement has as its goal the global regulation of violent extremism on the Internet and in social media messaging. Ardern calls upon assembled representatives of Facebook, Google, and Twitter to lead the way towards an online world that is both free and harm-free by enforcing their existing standards and policies about violent and racist content, improving response times involved in removing such content when it is reported, removing accounts responsible for posting content that violates the platform’s standards, making transparent the algorithms that lead searchers to extremist content, and committing to verifiable and measurable reporting of their regulatory efforts. Affirming that the ability to access the Internet is a benefit for all, she also asserts that people experience serious harm when exposed to terrorist and extremist content online, and that we have a right to be shielded from violent hatred and abuse.

Why is this call different from all other calls?

What action can we expect in the wake of this call? And what consequences might plausibly flow from that action?

The Internet as a site of racist hate speech and vicious verbal abuse is not a revelation; in recent years, many culture-watchers and technology journalists have documented an increasingly bold and increasingly globalised “community” of white supremacists whose initial - sometimes accidental - radicalisation is reinforced in the echo chambers of this so-called “dark web”, the encrypted social messaging platforms that Ardern identifies as in need of regulation. (I put the word “community” in quotes here because the meaning derived from the word’s Latin root [*munis/muneris*: the word for *gift*] makes it a darkly ironic way to describe these bands of people: if *community* is a gift we share with each other, their gift of poisonous hate is one that damages all those with whom it is shared.

Recognising the danger of these groups, as Ardern does, and seeking to neutralise their effects on our online and in-person worlds is important, even urgent. As Syracuse University professor Whitney Phillips observes: “It’s not that one of our systems is broken; it’s not even that all of our systems are broken...It’s that all of our systems are working...towards the spread of polluted information and the undermining of democratic participation.”

The Internet as a site of racist hate speech and vicious verbal abuse is not a revelation; in recent years, many culture-watchers and technology journalists have documented an increasingly bold and increasingly globalised “community” of white supremacists whose initial - sometimes accidental - radicalisation is reinforced in the echo chambers of this so-called “dark web”...

The consequences of the way these systems are working are now as clear to New Zealanders in the wake of the Christchurch attacks as they have been to Americans, to Kenyans, to Pakistanis, and to Sri Lankans in the wake of their respective experiences of hate-fuelled terrorism. American Holocaust scholar Deborah Lipstadt reminds us that acts of violent hatred always begin with words, words that normalise and seek to justify the genocides, pogroms, and terror attacks to come. If we do not speak out against those words, she notes, we embolden the speakers in their drive to turn defamatory words into deadly actions.

So the action called for at Ardern and Macron's Christchurch summit is warranted. Will it happen? Will the nations who have the ability to exert moral pressure on the companies that created and profit from these online platforms actually force a change in how white supremacist rhetoric is dealt with? Karen Kornbluh, a senior fellow at the German Marshall Fund, who is quoted in [Audrey Wilson's May 15 Foreign Policy Morning Brief](#), thinks that "the best case scenario [for] the Call to Action provides the political pressure and support for platforms to increase vigilance in enforcing their terms of service against violent white supremacist networks."

The problem with reliance on political pressure to change cultural policies driven by economic incentives and reinforced by jurisdictional divides is that when the pressure fades, the behaviour we want changed re-emerges. This has certainly been the case in prior efforts to alter Facebook's inconsistent oversight of its users. Back in 2015, for instance, Germany's then Federal Minister of Justice and Consumer Protection, Heiko Maas, filed a written complaint with Facebook about its practice of ignoring its own stated standards and policies for dealing with racist posts. Maas pointed out the speed with which Facebook removes photographs (like those posted by breast cancer and mastectomy survivors who seek to destigmatise their bodies) as violations of the platform's community standards, and the corresponding inattention to user complaints about racist hate speech. A [Foreign Policy analysis](#) of Maas's complaint letter reports that it led to an agreement between German officials and representatives of Facebook, Google, and Twitter - the very same companies who sent representatives to Ardern and Macron's Christchurch summit - on a voluntary code of conduct that included a commitment to more timely removal of hate-filled content. That was in 2015; in Maas's view, Facebook has subsequently failed to honour the agreement.

The problem with reliance on political pressure to change cultural policies driven by economic incentives and reinforced by jurisdictional divides is that when the pressure fades, the behaviour we want changed re-emerges. This has certainly been the case in prior efforts to alter Facebook's inconsistent oversight of its users.

Even at the international/multi-national level at which Ardern's call is framed, it is not clear how much capability there is to reform the discursive violence inflicted on us by white supremacist digital hate cultures. Audrey Wilson's *May 15 Foreign Policy Morning Brief* reports that in the wake of his own visit to the Christchurch mosques that were the scene of white supremacist terror, UN Secretary-General António Guterres committed himself to combatting hate speech.

However, in a talk at the United Nations University in Shibuya (Tokyo) on March 26, 2019, Mike Smith, former Executive Director of the United Nations Counter-Terrorism Committee Executive Directorate, was pessimistic about the possibilities for monitoring sites on which people like the Christchurch killer engage in their mutual radicalisation. One could argue with some plausibility that the "soft power" of moral authority, widely acknowledged as one of the UN's key strengths, should be used to speak out against hate and terror lest its silence on the matter foster a sense of impotence on the part of the international community. However, as Smith made clear, that level of monitoring on the part of international institutions (or national ones for that matter) is not feasible,

even assuming there is no other claim on the resources that would be required. The only workable way to implement monitoring of online hate groups is for the tech companies to be doing it themselves and, as Ardern asked for in her Christchurch Call, to be reporting regularly on their efforts to international and national agencies.

What could possibly go wrong?

In considering the question of whether the Christchurch Call does, or can, mark the moment when the world begins to take white supremacist hate speech seriously, we need to consider what we are dealing with in that speech, in that “community”. One American [think-piece](#) published in the days following the Christchurch attacks observed that “[r]acism is America’s native form of fascism”, and I think it might be instructive to take that claim seriously. Frequently a carelessly-used and controversial epithet, fascism has been broadly defined as a political worldview in which some of a nation’s people have been given status as persons, as citizens, as lives that matter in a moral hierarchy, and others have had that status denied to them.

Seeing racism as a variant of fascism gives us the resources to understand why online white supremacist hate speech is such an intractable problem. Essayist Natasha Lennard, a theorist of the Occupy movement that erupted in the United States in 2011, insists that “fascism is not a position that is reasoned into; it is a set of perverted desires and tendencies that cannot be broken with reason alone.” Instead, [she argues](#) that fascism—which she defines as “far-right, racist nationalism”—must be fought militantly: white supremacists must be exposed, and the inadequately regulated online spaces where their views are promulgated must be shut down. A similar no-tolerance approach to the more mainstream sympathiser sites where these views are legitimised is also warranted as part of anti-fascist (antifa) organising, she thinks. The goal of those who oppose fascism, racism, and white supremacy must be to vociferously reject these views as utterly unacceptable.

The kind of intransigent approach Lennard advocates is precisely the posture that the companies providing these online platforms are so ill-equipped and unwilling to adopt. As *Foreign Policy* [writers Christina Larson and Bharath Ganesh](#) both make clear, social media platforms like Facebook have long cloaked themselves in a rhetoric of utopian connectedness and free speech. Absence of regulation has been pitched to users as the precondition of popular empowerment.

Ganesh points out that there is a real disparity of treatment in the ways online platforms deal with extremist speech: where German minister Heiko Maas charged that Facebook censors photographs involving nudity and leaves hate speech to flourish, Ganesh qualifies that only some speech is left unregulated. Extremist white supremacist hate speech is routinely ignored or approached with caution and with charitable concern for the poster’s rights of expression, but extremist jihadi speech is monitored, removed, and blocked. “There is a widespread consensus that the free speech implications of such shutdowns are dwarfed by the need to keep jihadi ideology out of the public sphere,” Ganesh explains. But, “right-wing extremism, white supremacy, and white nationalism...are defended on free speech grounds.”

In part, this is precisely because of the existence of more mainstream sympathiser sites (such as Breitbart, Fox, InfoWars) that ally themselves with right-wing politicians and voters, and defend white supremacists through “dog whistles” (key words and phrases that are meaningful to members of an in-group and innocuous to those on the outside), such that, as Ganesh puts it, this particular “digital hate culture...now exists in a gray area between legitimacy and extremism”. Fearing backlashes, howls of protest about censorship, and reduced revenue streams if users migrate out of their platforms, social media companies have consistently chosen to prioritise these users over the less powerful, less mobilised minority cultures who are undermined by digital hate.

Extremist white supremacist hate speech is routinely ignored or approached with caution and with charitable concern for the poster's rights of expression, but extremist jihadi speech is monitored, removed, and blocked.

In light of this self-serving refusal to apply their own community standards even-handedly, what we are likely to see from social media platforms in response to the Christchurch Call is more legitimising of white supremacy rhetoric that is increasingly entering the mainstream of American discourse, and more policing of already marginalized viewpoints and voices. The most likely result is of their caretaking of this current situation is proliferation of the inconsistent censorship Ganesh identifies, and extension of that censorship to the very groups and users who might be calling out white supremacy. [One example](#) of this censorship of anti-racism predating the Christchurch Call involved a group of feminist activists calling themselves "Resisters," who created an event page on Facebook to promote a 2018 anti-racism rally they planned for the anniversary of the Unite the Right hate rally in Charlottesville. Facebook removed the page on the grounds that it bore a resemblance to fake accounts they believed to be part of Russian disinformation efforts aimed at influencing the 2018 US mid-term elections.

What then must we do?

"The real problem is how to police digital hate culture as a whole and to develop the political consensus needed to disrupt it," Ganesh tells us. In his view, the central question of this debate about online hate is: "Does the entitlement to free speech outweigh the harms that hateful speech and extreme ideologies cause on their targets?" That question is also posed in the Christchurch Call, and in abstraction it is a difficult one. People committed to freedom and to flourishing social worlds want both the right to express themselves and protections against the violence and dehumanisation that hate speech enacts.

Practically speaking, however, we often can draw lines that delineate hate speech from speech that needs to be protected by guarantees of right of expression (often, views from marginalised communities). Ganesh cites Section 130 of the German Criminal Code as an example: in free, democratic Germany, it is nonetheless a criminal offense to engage in anti-Semitic hate speech and Holocaust denial. The point of this legal prohibition is to disrupt efforts to attack the dignity of marginalised individuals and cultures, which is, Ganesh contends, "what digital hate culture is designed to do." If our legal remedies begin - as the Christchurch Call asks all remedies to - with basic human rights and basic human dignity as their central concerns, they will not, he thinks, contravene our entitlement to express ourselves.

"The real problem is how to police digital hate culture as a whole and to develop the political consensus needed to disrupt it," Ganesh tells us. In his view, the central question of this debate about online hate is: "Does the entitlement to free speech outweigh the harms that hateful speech and extreme ideologies cause on their targets?"

Those who fear that any attempt to delineate speech undeserving of protection will slide down a slippery slope into censorship often turn for support to nineteenth-century British philosopher John Stewart Mill's impassioned argument for the necessity of robustly free speech in his 1859 work *On Liberty*. However, Mill's motivation for that argument was his belief that freedom of expression is a key component of human dignity. Free speech does have limits, even for Mill; he articulates those limits in arguably his most famous contribution to Western political theory: the harm principle, which says that limits on an individual's freedom are only justified to the extent that they prevent

harm to others.

Recognising that words have the capacity to trigger action, Mill acknowledges that a society cannot tolerate as protected speech a polemic to an angry mob outside the house of a corn dealer in which one charges the corn dealer with profiteering at the expense of hungry children and calls for death to corn dealers. Building on this view that incitement to reasonably foreseeable harm or violence warrants restrictions on speech, even the United States, with its expansive constitutional protections for speech, has enshrined limitations. (One cannot yell “fire” in a crowded theatre, for instance.)

While laws – and responsible oversight by social media platforms, if ever that can be mandated in ways they will adhere to – can structure the playing field, they cannot determine the actions of the players. For that necessary change, we must look to our own behaviours and attitudes and how each of us might play our role in reinforcing social norms. In a post-Christchurch attacks [interview](#), American anti-racist educator Tim Wise advises people: “Pick a side. Make sure that every person in your life knows what side that is. Make sure your neighbors know. Make sure the other parents where your kids go to school know what side you are on. Make sure your classmates know. Make sure that your family knows what side you are on. Come out and make it clear that fighting racism and fascism are central to everything that you believe.”

We must, I think, resist the temptation of the easy neoliberal “solution,” the fiction that small numbers of committed individuals can neutralise a normalised culture of hate. But there is a germ of insight in Wise’s prescription. Yes, we need a better legal climate, one that levies real penalties on social media platforms that fail to monitor the content they make available in our lives; yes, we need more responsible social media companies and Internet site moderators; and we also need to all do what we can to make sure that the people who are listening to each of us are hearing messages that contribute to a healthy and caring social world.

One thing I learned from the 2014 online frenzy of misogynist hate known as “GamerGate” (the campaign of invective and abuse organised against women in the video game industry) was that a small number of committed individuals can *produce* a normalised culture of hate. Another thing I learned was that many of the casual reproducers of that organised hate are not fully culpable actors; they have been drawn into something they think they understand but when they can be made to see how harmful it is, they will renounce it. I do think Natasha Lennard is right about the futility of trying to appeal to people who have chosen hate or fascism, but there are many others on the fringes who can be influenced away from those ideas. They need to be surrounded by people in their (online and offline) lives who are speaking the language of anti-racism, feminism, multicultural inclusion, and the equal right to dignity of all human beings.

One thing I learned from the 2014 online frenzy of misogynist hate known as “GamerGate” was that a small number of committed individuals can *produce* a normalised culture of hate.

If online hate has IRL (in real life) ramifications, then IRL influencing might be a way to save or reclaim some otherwise radicalised young people, and also a way to assert pressure on the social media platforms to “walk their talk” of wanting a more connected community. The Christchurch Call cannot, in and of itself, drive out the poison of white supremacist hate. But it can, perhaps, inspire us to make our communities (the gifts we share with each other) gifts worth receiving.

Published by the good folks at [The Elephant](#).

The Elephant is a platform for engaging citizens to reflect, re-member and re-envision their society by interrogating the past, the present, to fashion a future.

Follow us on [Twitter](#).

