Journal & Proceedings of the Royal Society of New South Wales, vol. 151, part 1, 2018, pp. 69–81. ISSN 0035-9173/18/010069-13

Algorithms of hate: How the Internet facilitates the spread of racism and how public policy might help stem the impact

Andrew Jakubowicz

Social and Political Sciences, University of Technology Sydney, Sydney, Australia

E-mail: A.Jakubowicz@uts.edu.au

Abstract

Complex multicultural societies hold together through effective and interactive communication, which reinforces civility, enhances information sharing, and facilitates the expression of interests while permitting both diversity and commonality. While trust is an important cement in the building of social cohesion, multicultural societies face continuing challenges as their ever-extending populations test the trust necessary to constitute supportive, bridging social capital. The Internet, which has become a crucial component of the communication systems in modern societies, offers both opportunities and challenges, especially in the generation and circulation of race hate speech which attacks social cohesion and aims to impose singular and exclusive racial, ethnic or religious social norms. The Internet in Australia remains problematic for four key reasons. The underlying algorithms that produce social media and underpin the profitability of the huge domains of Facebook and Alphabet also facilitate the spread of hate speech online. With very limited constraints on hate speech, the Australian Internet makes it easy to be racist. Human/computer interactions allow for far greater user disinhibition, which suits the proclivities of those more manipulative and sadistic users of the Internet. All of this is occurring in a post-truth world where racially, religiously and nationalistically inflected ideologies spread fairly much unchecked, and discourses of violence become everywhere more apparent. Australia has opportunities to do something about this situation in this country, yet we see around us a lethargy and acceptance of technological determinism. The paper assesses these claims and proposes some ways forward that are evidence-based, and collaborative, scholarly and social.

Post-truth and Internet racism: knowledge and power

The Forum on Post Truth organised by the Royal Society of NSW and the scholarly academies, held in November 2017, focuses our attention on the concept of truth, its meanings in the "hard" and social sciences, and the manipulation of public comprehension of the realities in which we live. As a sociologist with humanist tendencies I have long held that truth *claims* are just that: propositions that can be tested empirically. However what counts as evidence can more often be a question for vigorous debate, though simple assertion cannot win the day. We have seen in this Forum a variety of approaches to this issue, with particular focus on the interfaces between science and power, between scholarship and politics. Perhaps one of the most complex interchanges – between knowledge and prejudice, freedom and constraint, emotion and rationality, and policy and ideology – can be found in the rapidly burgeoning space of on-line racism.

On-line racism is a comparatively new phenomenon, maybe a generation old, given its dependence on the invention of the Internet and the development of the World Wide Web (Brown, 2017). Racism, of course, has a much longer timeframe, drifting back into the mists of pre-history. Together racism and the Internet have produced a phenomenon that requires a truly interdisciplinary scholarship to describe and analyse, drawing on physical, economic, political and social sciences. Beyond my analysis in this article lies a prognosis on the one hand, and suggested programs for intervention on the other. This paper draws on a larger collegial work (Jakubowicz et al., 2017a) to make some specific claims about the way in which on-line racism serves the purposes of the expansion of "post-truth". The Internet facilitates this expansion by feeding a societal discourse in which race is given a false scientific realism, racism confirmed as an acceptable mode of social relationship, and the politics of racial prejudice allowed to permeate arguments about appropriate public policy (Nicholas and Bliuc, 2016).

Why cyber racism matters

Modern Australia has been described as a multicultural society, the most successful in the world according to Prime Minister Malcolm Turnbull, a perspective only possible if the Indigenous presence in Australia is ignored (Jakubowicz, 2015). Whether Australia in fact stands first in line — and I dispute this claim even in relation to the cultural diversity of immigrant descendants: Canada is far ahead on many criteria (Tierney, 2007) — multicultural societies all depend on a pro-active building of trust between disparate peoples, usually prompted and promoted by government. Trust, often described as though it were the glue that anchors social cohesion (Markus, 2015), can be fragile in a multicultural milieu, where people do not go back many generations together, and the intimate ties of kin and communal sharing among strangers are less evident. Moreover, the subtleties of cultural participation and understanding take time to evolve and modify the emotional and intellectual portfolios people draw on to interact with others different from themselves. Thus multicultural societies require active interventions in the public sphere to build community and resolve conflicts (Kymlicka, 2007). With the advent of the Internet, digital technologies are now deeply implicated in nearly all spheres of social interaction.

The issue of cyber racism has particular relevance for scientists, humanists and policy makers, as the phenomenon depends on the state of the social relations of multicultural societies, public policy perceptions and responses to those relations, and the affordances of the digital technologies. It thus "pitches" at a point where the academies intersect, the world-views and technical skills of the different branches can be applied, and the social advancement that the Royal Society seeks to nurture is being challenged. On the other hand citizens might ask why Australian society should be concerned about the spread of race hate speech on the Internet (Bernardi, 2016). Surely, in a liberal democracy, freedom of speech, no matter how objectionable, must be defended as a higher-order value, one linked directly to the pursuit of truth and therefore an underpinning of science? While people may take offence at what other people say about them, so long as the language does not seek to trigger or actually triggers criminal behaviour, do we not all have an interest in allowing its free expression?

In answer to these questions, let me begin with a short personal anecdote. Late last year I wrote a piece for *The Conversation* reviewing the question of whether the concept of ethno-political hierarchy or ethnocracy (Jakubowicz, 2016) — used to examine how race, religion or creed was either actively or unconsciously reflected in structures of sectarianised democratic power — could be usefully applied to Australian multiculturalism. My argument was attacked by a post-truth advocate who alleged its thrust would erode the importance of White Anglo-culture as the underpinning of Australian moral order. In addition, the individual pointed me and other readers towards a website, twitter feed and Facebook page (Di Stefano and Esposito, 2016) in which my article and myself as its author were the primary targets. The authors of that piece had headed the article with a photo of the ceiling of the Yad Vashem memorial hall to the slaughtered of the Holocaust in Jerusalem, while the article attacked me as Jewish and therefore implacably fixated (it appeared to them) on a project to destroy White Australia by advancing multicultural ideas. There were many other subtle and not so subtle references to the benefits of Nazism and the appropriate end for a Jew, to which the ceiling image of thousands of dead referred.

It is one of the uncomfortable consequences of being a Jewish intellectual and social scientist in the era of post-truth that the new Nazis and other ultra-nationalists find us particularly attractive as targets, both for the views to which we can be attached, and as individuals who can be made to suffer emotionally through activation of Holocaust tropes. Ultimately, I decided to take no action other than use the intervention as a standing case study in how the Internet has allowed the resurgence of race hate and the difficulties the system creates for any action to seek either redress or removal in a sea of global anonymity.

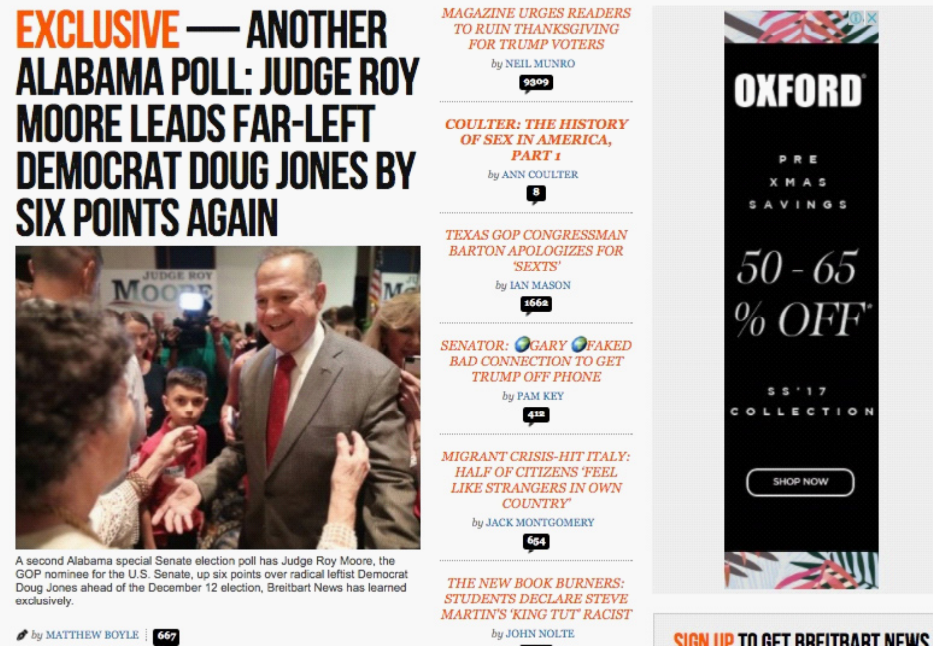
Four reasons Australia is a good place to be an online racist

Four main elements make the Australian experience of race hate on the Internet quite specific, though perhaps only slightly more intense or focused compared with its spread elsewhere. After all, the Internet has become a global network of interconnectivity, with instantaneous communication facilitating interactions between people who might in the past have never come into contact. This facilitation depends on both the physical/technological connections, and the technical languages and calculations that allow messages to flow and reach their targets. These algorithms or sets of rules have been layered over the short history of the Internet into vast portfolios of instructions, often requiring millions of calculations, with consequences both intentional and unintentional (Parish, 2017) (Buni and Chemaly, 2016).

The inventor of the World Wide Web, Tim Berners-Lee, has increasingly been worried by these unintended consequences. Early in 2017 he noted "And the thing that worries me most is that whatever it is we've created we've licensed racism to run free across the planet and the consequences of that for civilisation and democracy are very, very sordid if they're not addressed" (Berners-Lee, 2017). Near the end of 2017 he persisted with these concerns. "My vision for an open platform that allows anyone to share information, access opportunities and collaborate across geographical boundaries has been challenged by increasingly powerful digital gatekeepers whose algorithms can be weaponised by master manipulators" (Solon, 2017).

There are two sets of algorithms that are most implicated in this process, apart from the ones "weaponised" in spheres of civil contestation and those activated in "hot war" situations. Racism can be served either by directing Internet users to racist sites, or delivering racist messages to other sites. Both of these procedures are triggered by agglomerating data from multiple sources, and looking for patterns — patterns that are known to be profitable, though often cloaked in the language of enhancing user experience. The tie-in of the algorithms to the business models underpinning the Google empire (including YouTube) and Facebook makes them extremely difficult to change. In these circumstances, the platforms have been trying to find ways to limit the use of expensive human staff to monitor breaches of their user codes of conduct, while discovering that they have often been gamed by extremist Internet users and hackers who trip the faults deep inside the algorithmic hold-alls (Greenberg, 2016).

The specific interventions by extremists have both gender and class dimensions, as well as race. For example, the audiences most attuned to racist material in Western societies tend to be younger White males, a somewhat affluent category with disposable incomes, highly sought after by mainstream advertisers for products such as Coca-Cola and the UK military recruitment. Affluent males are also sought by media outlets such as *The Guardian*. These were the types of advertisers that in March 2017 found their messages appearing on racist, sexist and violent sites, and those associated with extremist White Power and Islamist organisations (young males not necessarily White). Many advertisers withdrew their campaigns from YouTube and Facebook, and tried to have Google change its ranking algorithms to avoid their placement in unacceptable locations during online searches (Statt, 2017) (Mostrous, 2017) (England, 2017).



EXCLUSIVE — ANOTHER ALABAMA POLL: JUDGE ROY MOORE LEADS FAR-LEFT DEMOCRAT DOUG JONES BY SIX POINTS AGAIN

A second Alabama special Senate election poll has Judge Roy Moore, the GOP nominee for the U.S. Senate, up six points over radical leftist Democrat Doug Jones ahead of the December 12 election, Breitbart News has learned exclusively.

by MATTHEW BOYLE 667

MAGAZINE URGES READERS TO RUIN THANKSGIVING FOR TRUMP VOTERS
by NEIL MUNRO
9309

COULTER: THE HISTORY OF SEX IN AMERICA, PART 1
by ANN COULTER
8

TEXAS GOP CONGRESSMAN BARTON APOLOGIZES FOR 'SEXTS'
by IAN MASON
1662

SENATOR: GARY FAKED BAD CONNECTION TO GET TRUMP OFF PHONE
by PAM KEY
412

MIGRANT CRISIS-HIT ITALY: HALF OF CITIZENS 'FEEL LIKE STRANGERS IN OWN COUNTRY'
by JACK MONTGOMERY
654

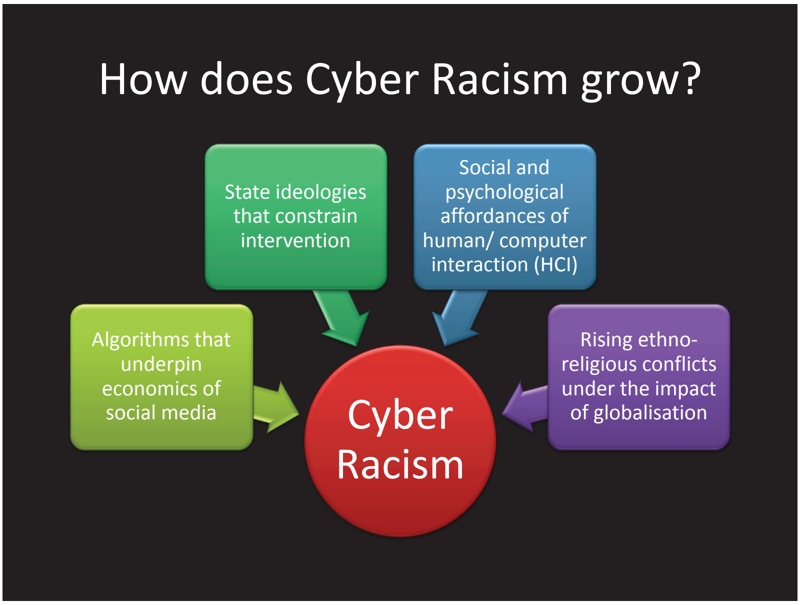
THE NEW BOOK BURNERS: STUDENTS DECLARE STEVE MARTIN'S 'KING TUT' RACIST
by JOHN NOLTE

Figure 1.

Readers can try this experiment themselves as I did. I am an occasional customer for a well-known men's clothing brand; I buy in-store although the company has my email for marketing purposes. When I searched the U.S. White Power Breitbart site for information using Google and Chrome, I was served advertisements for that clothing brand (see Figure 1). I also received arthritis treatment information, suggesting that Chrome had been logging my online therapy visits following my recent knee replacement operation. Both advertisements relate to White males: both Breitbart and the arthritis pill target older White males amongst their primary targets. Breitbart was intent to increase the spread of alt-right post-truth and pro-White Power discourses among its visitors, a process that both the advertiser and Alphabet were facilitating and helping to fund (through click-through visit payments where these occurred) (Amend and Morgan, 2017) (Anglin, 2016).

In other situations, algorithms may learn or be programmed to exclude people of colour from access to more highly valued user experiences. A review article in *Science* recently reported how "machine learning of semantics automatically shapes itself to human biases in language, in terms of race gender and disability" (Caliskan et al., 2017). In another instance, some facial recognition software cannot "read" the faces of people of colour and thus excludes them or their responses. In discussing these instances, the U.S.-based advocacy group, the Algorithmic Justice League, conceptualises the issue as "the bias of the coded gaze" (Algorithmic Justice League, 2016, Buolamwini, 2016).

Facebook has been alleged to have been involved in "multicultural affinity targeted advertising" by offering redlining algorithms that identify people on the basis of their race and restrict their access to offers of housing, employment or loans, thereby segmenting markets and populations into those who are



How does Cyber Racism grow?

State ideologies that constrain intervention

Social and psychological affordances of human/ computer interaction (HCI)

Algorithms that underpin economics of social media

Cyber Racism

Rising ethno-religious conflicts under the impact of globalisation

Figure 2.

acceptable and exploitable, and those who are rendered unacceptable and discardable. (Chaykowski, 2016)

The four factors that contribute to the extent and composition of cyber racism in specific jurisdictions can be summarised through the four "feeders" portrayed in Figure 2. All four are necessary to allow cyber racism to flourish, although the extent of each may vary across the globe. However, global, national, scientific and individual factors all play a role, while political action can have some impact on raising or reducing the "volume" of each parameter.

Racism on line

Racism has a long and controversial relationship to science. In 1875 Charles Darwin wrote that, as the science of humanity improves, so then human kind (and especially his peers of white European men of wealth and social status) would be drawn to "extend our sympathies to all men" (Paul, 1988). However, we know the actual trajectory of human history drew exactly the opposite perspective, creating from Darwin's insights the most cruel and vicious separations between peoples. The differences that Darwin saw within humanity became hierarchies of superordination in the ideologies of racism, where the empires of his time drew on poorly understood "truths" to generate overwhelming technologies of destruction. If "race" in all its manifestations finally proved to be an unacceptable framework for building human societies, it did not depart human consciousness at the end of the Second World War.

When UNESCO in 1950 first sought to deal with the science of race, it concluded that races were real categories of differentiation, though quite "inter-breedable" (UNESCO, 1950/1954/1957/1969). At the time the empires of the European centuries of expansion had not quite dissolved and their subordinated racially-justified colonial subjects had not yet reached independence. When the UN once more addressed what racism was in 1967, the world had changed. A global convention against racial discrimination had been passed, any notion that race had a scientific meaning had been abandoned, with UNESCO concluding "Racism stultifies the development of those who suffer from it, perverts those who apply it, divides nations within themselves, aggravates international conflict and threatens world peace" (UNESCO, 1950/1954/1957/1969).

If we take this to be a widely verifiable truth about the effects of racism, then the next factor that effects the extent and nature of racism in Australia lies in the ideologies that are expressed through legislation and action by the state that might follow such laws (McGonagle, 2012). Unlike many other countries, race hate speech is not criminalised in Australia at the national level. Indeed, Australia shares with the USA the reality that one can say anything about other ethnic and racial groups up to the point where advocacy of a crime or of violence is expressed. Australia in 1966 followed the lead of the US (Harris, 2008) to include a reservation in its ratification to article 4a of the International Convention on the Elimination of All Forms of Racial Discrimination (ICERD). Throughout 2016 the Federal Government sought for a second time in recent memory (previously in 2013/2014) to reduce the coverage of the Racial Vilification provisions of the Racial Discrimination Act, which had been introduced (as Section 18c) in 1976 (Baxendale, 2017).

While both those moves failed, the clear message from government was that racial vilification would be an acceptable practice to be defended behind the rubric of freedom of speech. However research by the CRaCR team (Jakubowicz et al., 2017b, Jakubowicz, 2017) in 2013 and 2016 demonstrated that only a small minority of Australians wanted there to be unrestricted freedom to vilify people on the basis of their race or ethnicity (Parliament of Australia, 2017). Even so, Australia remains one of the easiest places to be racist online (Hunyor, 2008), providing only slow and difficult systems to file complaints, a reluctant and resistant set of corporate providers of Internet services, and a confusing and overlapping set of regulatory regimes.

All these interactions take place in a global environment of heightened fear and tension associated with distinctions based on ideas about race, religion, ethnicity and nationhood. In large part these tensions have grown far beyond the earlier penetration of such issues during epochs dominated by print, audio or even television communication, because of the omnipresence of the Internet and the surge of post-truth propaganda and dissimulation. Thus, the technology and the circumstances have interacted and exponentially expanded the impact of hatred on fearful communities. Over the past decade or more, such divisions have become normalised in stereotypical and increasingly hostile and hurtful encounters, the veracity of which has become impossible to test.

The Internet depends on the easy anonymity of its users, the effective asynchronicity of its interactions, and the isolated circumstances under which most people engage with others online. Such human/computer/human interaction allows for social and psychological opportunities that would be far more difficult in the everyday world. So using the Internet intensifies "disinhibition" (Martin, 2013) (Suler, 2004), by allowing sadistic, egoistic and manipulative behaviour to spread more fluidly (Brown, 2017, Stein, 2016). There is considerable evidence that such dynamics are reflected in the small number of people who apparently "produce" racism online, with a large number of people encountering it, in its many forms, as bystanders.

Extent of racism	**Online:**	**Target**	**Perpetrator**	**Bystander**
Opponent of racism/ not prejudiced		2.2% Often seen as carrying responsibility; opposes racism online; defends from attack.	0.7% Asserts own ethno-religious group superiority while decrying racism.	15.3% Once alerted to issues, becomes more aware; often seen as main bulwark against racism.
Unconcerned/ moderately prejudiced		7.9% Alerted to racism when targeted; tends to withdraw from exposure.	2.8% Unaware amplifier; likes racist joking etc; drawn to swarm.	63.1% Doesn't recognise or withdraws from exposure; can be unaware supporter.
Proponent of racism/ strongly prejudiced		1% Activist responder engaged in fight with perceived harassers.	3.7% Sharp end of racist propaganda; seeks to build following; advance racist agendas.	5.7% Lurks to like; aware amplifier; not pro-active but strong supporter.

Table 1. Algorithms of Hate tables etc.

A 2000+N online survey undertaken for the CRaCR project in 2013 (Jakubowicz et al., 2017a, ch.3) provided data for an exploration of the relationship between the range or type of encounters online, and the attitudes of the subjects on issues associated with racism. Six items from a 20+ compilation of items eliciting responses to attitudes on ethnic and cultural differences on a seven-point original scale, provided a three-point scale of attitudes. Target, Perpetrator and Bystander were discrete categories, although a few individuals were in two or all.

From this distribution, it is possible to have a sense of how different users of the Internet, based on their own levels of prejudice, deal with encounters with racism. The picture is quite complex, demonstrating the interactive nature of the web and the changing position of people who are activist. About 10% of Targets show high levels of prejudice while most Targets show little (71%) or none. Over 50% of Perpetrators are strongly prejudiced, while only 10% show no signs of prejudice. The largest group in relation to racism by far are Bystanders, who make up over 80% of Internet users. Of their number about 7% are highly prejudiced, about 75% moderately so, and 18% show very low levels of prejudice.

The distinctions, based on the level of prejudice and type of encounter, point to the online activities associated with each category, and thereby, what policy and practice responses may be appropriate. These are summarised in each cell. The dynamic of the Internet world of race hate becomes evident — users are making decisions, engaging or withdrawing, being harassed or harassing, in a constantly moving environment. For the Perpetrators one of the goals is to "game" those defences that platforms provide, while seeking to normalise hate speech and thereby transform the social relations of the Internet into one infused with racist ideology and perspectives. Each Internet user category is positioned in specific ways in relation to the expansion of online racism.

However, Targets are often expected to carry the burden of response, or are abandoned to that fate. In the Australian context agencies such as MulticulturalNSW have been charged by their political managers in recent times with implementing an anti-racism/pro-multicultural agenda online; however, these can easily be wound back under ministerial direction should ideologies change and predilections for addressing racism become less pressing. In the federal sphere there have been no such agencies, as political attacks on the Australian Human Rights Commission from the Government have limited its capacities to do so. However, the AHRC has been active in the Racism It Stops With Me campaign, and associated online and broadcast advertisements about racism. Even so the Commission cannot intervene in the online world without direct complaints to pursue. However, the Children's E Safety Commissioner has begun to initiate workshops and strategies to build capacity among threatened communities to defend themselves and advance alternative "truths" against racist attacks.

Increasingly, Bystanders are recognised as extremely important potential defenders of Targets and crucial participants in pushing back against racist hate speech (Nelson et al., 2012). Given that racists want to ensure that every space they enter becomes infected and then permeated by their ideology and discourses, resisting such entry-ism and denying racists these local victories, however appalling, cruel and foul their language, contributes to a more open Internet.

Strong proponents of racism who are Perpetrators make up less than 4% of users, yet they generate and stimulate the vast array of hate speech in its text, meme and video forms. They are supported by a larger group of Bystanders, who "lurk to like", and want to extend the reach of their swarm leaders into the moderately prejudiced huge bulk of Bystanders. Their attachment to such discourses is closely associated with their belief that their views are widely shared, a position reinforced when they find the sites they like carry no opposition messages or signs of antiracist arguments.

The complexity of the field indicates the need for more coherent and science-based policy; government and civil society interventions in such situations would help reassert both the value of truth, and the right to a democratic and civil Internet (Daniels, 2010). Without an Internet in which truth can be asserted and demonstrated, the overall edifice of evidence-based argument and policy continues to crumble, and issues far removed from racism are caught in a wave of beliefs in which truth and science have no hold (Miller, 2016).

We can summarise the current nexus in Australia through these CRaCR project findings. The basic technologies underlying the spread of race-hate filled social media and related technologies are not easily amenable to state action, especially where the algorithms are so rooted in the profitability of the platforms. Governments fail to realise how much the social cohesion they promote constantly faces attacks that seek to unwind the trust and social capital upon which it depends.

The bad behaviour that promotes the spread of race hate can be quickly and widely replicated (Phillips, 2015). In the process the Internet emotionally and often financially rewards the dark triad behaviour of narcissism, manipulation, and lack of empathy (Binns, 2012). The Perpetrators gain emotional reinforcement, a sense of purpose, and a continuing stream of supportive followers when they are left unchecked and unrestricted; even more so when they are morally castigated but effectively allowed to continue unconstrained. Yet, for anti-racists, taking on the Perpetrators and inventively resisting racist hate speech remains a challenging and wearying activity, with little of the emotional reinforcement that sustains and rewards the Perpetrators (Gagliardone et al., 2015).

The resistance to racism can be further weakened where political leaders are averse to taking courageous positions on difficult issues, being more likely to be drawn to the pressures from conservative post-truth groups that they celebrate freedoms rather than constraints (Group of Eminent Persons of the Council of Europe, 2011).

The major corporations such as Facebook and the Alphabet stable (Google, Facebook, Instagram etc.) appear more interested in protecting their economic interests than in resolving the questions of social impact generated by their business models (Levine, 2013) (Zuckerberg, 2017). For example, they are reluctant to expose themselves to critical scholarly research. They will respond to Parliamentary interrogation, however, when they perceive their interests may be served (Garlick, 2017, Garlick, 2018). Despite widespread criticism by organisations such as the Simon Wiesenthal Institute in the USA (Simon Wiesenthal Center, 2017) and the Online Hate Prevention Institute in Australia (Online Hate Prevention Institute, 2015), the two great Internet behemoths have gone to great lengths to protect their underlying business models from changes that might be thought necessary by critics to address the pervasiveness of racism throughout their services.

In 2017 and into 2018 the Australian Senate Standing Committee on Legal and Constitutional Affairs undertook an examination of the adequacy of Australian criminal offences in relation to cyber-bullying. While racist hate-speech is often part of cyber-bullying, it is far less likely to attract attention than do other dimensions of bullying and harassment. Facebook made two written submissions in addition to its oral evidence. In the first (Garlick, 2017) the company stressed that platforms should be excused from any responsibility for material published in their pages by their users, as the company was not a publisher in the traditional pre-Internet sense, and that it already responded quickly to requests from affected parties, or the police, for bullying material to be taken down. In a return submission responding to questions on notice from the Committee, the company representative described the strategies adopted to deal with complaints and problematic users: Facebook noted it had 14,000 people working worldwide on community operations in 2017, and was planning to increase this number to 20,000 in 2018.

That is, one key area was human intervention, leaving the fundamental algorithms tweaked but not significantly changed. Discussing Facebook's "removal of hateful content in Europe," the company pointed to the agreement between social media firms and the European Commission to tackle the "problem of hate speech in Europe". Pushing back against the German law that criminalises activities of companies that fail to meet take-down standards, Facebook believed that "There is no place on Facebook for hate speech … industry codes are a more collaborative and effective [way] of achieving the results we all want to see" (Garlick, 2018). Australia has nothing like the European Commission Code of Conduct; Facebook made no offer that they would collaborate with civil society and government to ensure that one could be established.

Building resilience

However, associations that bring together people concerned with both civility and truth do have avenues open for them. They can be part of the move to build civil society alliances that abhor racism, and seek to push back against the acceptance or legitimisation of racism and racist discourses. Where initiatives in the legal sphere are opening, as a consequence for instance of the decision of the E Safety commissioner to recognise racism as a problem, then innovations such as the New Zealand Harmful Digital Communications Act could be considered (New Zealand Law Commission, 2012). The Australian Human Rights Commission could be both permitted and resourced to identify and pursue particularly egregious cases of cyber racism where no Target would otherwise be prepared to come forward. Civil society groups could call out and publicise, through social media, advertisers who allow their names to be associated with race hate sites, thus putting pressure on the large platform providers to find strategies to reduce such associations.

Perhaps the Royal Society and the Academies, with their aspirations to link science with human prosperity and well-being, might well take on strategy development that looks to public policy based on science as a way forward (Came and Griffith, 2017). A small group of mathematicians, philosophers, social scientists and others might workshop such ideas to contribute to crowd-sourcing resilience strategies, so that the algorithms that underpin social media in the future are not so conducive to the proliferation of hate: indeed, algorithms if not of love then at least of peace might eventuate. Ultimately resilience requires strong networks that build active cells of knowledge, where racism can find no place to flourish.

References

Algorithmic Justice League (2016) *The Coded Gaze* https://www.ajlunited.org/the-coded-gaze 2018).

Amend, A. and Morgan, J. (2017) Breitbart under Bannon: Breitbart's comment section reflects alt-right, anti-Semitic language, *Hatewatch*. https://www.splcenter.org/hatewatch/2017/02/21/breitbart-under-bannon-breitbarts-comment-section-reflects-alt-right-anti-semitic-language (Accessed 23 February 2017).

Anglin, A. (2016) A normie's guide to the alt-right, *The Daily Stormer*. http://www.dailystormer.com/a-normies-guide-to-the-alt-right/ (Accessed 15 January 2017).

Baxendale, R. (2017) Ethnic leaders united against substantive changes to section 18C, *The Australian*, 22 March. http://www.theaustralian.com.au/national-affairs/ethnic-leaders-united-against-substantive-changes-to-section-18c/news-story/534963bdb7dc4346bc75dbafad7fb688.

Racial Discrimination Amendment Bill 2016 Explanatory Memorandum.

Berners-Lee, T. (2017) Tim Berners-Lee: I invented the web. Here are three things we need to change to save it, *The Guardian*. https://www.theguardian.com/technology/2017/mar/11/tim-berners-lee-web-inventor-save-internet (Accessed 15 March).

Binns, A. (2012) Don't feed the Trolls! Managing troublemakers in magazines' online communities, *Journalism Practice,* 6(4), pp. 547-562.

Brown, A. (2017) What is so special about online (as compared to offline) hate speech?, *Ethnicities*, Sage. http://journals.sagepub.com/doi/10.1177/1468796817709846 (Accessed 3 June 2017).

Buni, C. and Chemaly, S. (2016) The secret rules of the Internet, *The Verge*. http://www.theverge.com/2016/4/13/11387934/internet-moderator-history-youtube-facebook-reddit-censorship-free-speech.

Buolamwini, J. (2016) InCoding—in the beginning, *Medium.com*. https://medium.com/mit-media-lab/incoding-in-the-beginning-4e2a5c51a45d.

Caliskan, A., Brysin, J. and Narayan, A. (2017) Semantics derived automatically from language corpora contain human-like biases, *Science*, 356(6334). http://science.sciencemag.org/content/356/6334/183.full.

Came, H. and Griffith, D. (2017) Tackling racism as a "wicked" public health problem: enabling allies in anti-racism praxis, *Social Science and Medicine*, Elsevier. http://dx.doi.org/10.1016/j.socscimed.2017.03.028.

Chaykowski, K. (2016) Facebook to ban "ethnic affinity" targeting for housing, employment, credit-related ads, *Forbes/Tech/#Getting Buzz*.

Daniels, J. (2010) Cyber racism & the future of free speech (updated), *Racism Review: scholarship and activism towards racial justice*, 2016(13 November). http://www.racismreview.com/blog/2010/11/16/cyber-racism-future-of-free-speech/ (Accessed 16 December 2016).

Di Stefano, M. and Esposito, B. (2016) Australia has an alt-right movement and it's called #DingoTwitter: the white nationalists are down under, *Buzzfeed News*. https://www.buzzfeed.com/markdisefano/gday-pepe (Accessed 4 May 2017).

England, C. (2017) Pewdiepie dropped by Disney over YouTube star's anti-Semitic videos, *Independent* 2017].

Gagliardone, I., Gal, D., Alves, T. and Martinez, G. (2015) *Countering Online Hate Speech* Paris: UNESCO. http://unesdoc.

unesco.org/images/0023/002332/233231e.pdf (Accessed: 21 February 2017).

Garlick, M. (2017) Submission 4 by Facebook to Senate Constitutional and Legal Affairs References Committee, *Reference: Adequacy of existing offences in the Commonwealth Criminal Code and of state and territory criminal laws to capture cyberbullying*. https://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Legal_and_Constitutional_Affairs/Cyberbullying/Submissions.

Garlick, M. (2018) Facebook and Instagram answers to questions on notice, Senate Constitutional and Legal Affairs References Committee, *Reference: Adequacy of existing offences in the Commonwealth Criminal Code and of state and territory criminal laws to capture cyberbullying*.

Greenberg, A. (2016) Inside Google's Internet Justice League and its AI-powered war on trolls, *Wired,* October.

Group of Eminent Persons of the Council of Europe (2011) *Living together: Combining diversity and freedom in 21st-century Europe* Council of Europe.

Harris, H. (2008) Race across borders: the U.S. and ICERD, *Harvard BlackLetter Law Journal*, 24, pp. 61-67. http://hjrej.com/wp-content/uploads/2012/11/vol24/Harris.pdf (Accessed 20 April 2016).

Hunyor, J. (2008) Cyber-racism: can the RDA prevent it? *Law Society Journal* May, pp. 34-35.

Jakubowicz, A. (2015) "Hating to know": government and social policy research in multicultural Australia, in Husband, C. (ed.) *Research and Policy in Ethnic Relations*. Bristol: Policy Press, pp. 53-78.

Jakubowicz, A. (2016) First the word, then the deed: how an "ethnocracy" like Australia works, *The Conversation*. https://theconversation.com/first-the-word-then-the-deed-how-an-ethnocracy-like-australia-works-69972 (Accessed 12 March 2017).

Jakubowicz, A. (2017) What did Galaxy's poll tell us about freedom of speech and 18C? Not what the IPA said it did, *The Conversation*. https://theconversation.com/what-did-galaxys-poll-tell-us-about-freedom-of-speech-and-18c-not-what-the-ipa-said-it-did-72197 (Accessed 3 February).

Jakubowicz, A., Dunn, K., Paradies, Y., Mason, G., Bliuc, A.-M., Bahfen, N., Atie, R., Connelly, K. and Oboler, A. (2017a) *Cyber Racism and Community Resilience: Strategies for Combating Online Race Hate* London: Palgrave Macmillan. http://www.palgrave.com/de/book/9783319643878 (Accessed: 5 October 2017).

Jakubowicz, A., Dunn, K. and Sharples, R. (2017b) Australians believe 18C protections should stay, *The Conversation*. https://theconversation.com/australians-believe-18c-protections-should-stay-73049 (Accessed 17 February).

Kymlicka, W. (2007) *Multicultural Odysseys: Navigating the New International Politics of Diversity.* Oxford: Oxford U. P.

Levine, M. (2013) Controversial, harmful and hateful speech on Facebook. https://www.facebook.com/notes/facebook-safety/controversial-harmful-and-hateful-speech-on-facebook/574430655911054/.

Markus, A. (2015) Mapping social cohesion: The Scanlon Foundation surveys 2015, Available: Scanlon Foundation, Australian Multicultural Foundation, Monash University. http://scanlonfoundation.org.au/wp-content/uploads/2015/10/2015-Mapping-Social-Cohesion-Report.pdf (Accessed 14 November 2016).

Martin, A. (2013) Online disinhibition and the psychology of trolling, *Wired*. http://www.wired.co.uk/article/online-aggression (Accessed 17 August 2016).

McGonagle, T. (2012) The troubled relationship between free speech and racist hate speech: the ambiguous roles of the media and internet, *Day of Thematic Discussion "Racist Hate Speech"*, Available: UN Committee on the Elimination of Racial Discrimiantion. https://www.ivir.nl/publicaties/download/Expert_Paper_Racist_Hate_Speech_UN.pdf (Accessed 24 May 2016).

Miller, T. (2016) Cybertarian flexibility—when prosumers join the cognitariat, All That Is Scholarship Melts into Air, in Curtin, M. & Sanson, K. (eds.) *Precarious Creativity: global* *media, local labor*. California: University of California Press.

Mostrous, A. (2017) YouTube hate preachers share screens with household names, *The Times* (17 March). http://www.thetimes.co.uk/article/youtube-hate-preachers-share-screens-with-household-names-kdmpmkkjk (Accessed 25 March 2017).

Nelson, J., Paradies, Y. and Dunn, K. (2012) Bystander anti-racism: a review of the literature, *Analyses of Social Issues and Public Policy*, Accepted 26.08.11.

New Zealand Law Commission (2012) Harmful Digital Communication: The adequacy of the current sanction and remedies, *Ministerial Briefing Paper, The New Zealand Law Commission, New Zealand Parliament*.

Nicholas, F. and Bliuc, A.-M. (2016) "It's okay to be racist": moral disengagement in online discussions of racist incidents in Australia, *Ethnic and Racial Studies,* 39(14), pp. 2545-2563.

Online Hate Prevention Institute (2015) Fight Against Hate, *OHPI*. https://fightagainsthate.com/ (Accessed 13 March 2016).

Parish, E. (2017) Mapping the geography of racism: why deep dives in data matter, *hastac*. https://www.hastac.org/blogs/erinparish/2017/04/24/mapping-geography-racism-why-deep-dives-data-matter-0 (Accessed 4 June 2017).

Parliament of Australia (2017) *Freedom of speech in Australia: Inquiry into the operation of Part IIA of the Racial Discrimination Act 1975 (Cth) and related procedures under the Australian Human Rights Commission Act 1986 (Cth)* Canberra: Parliament of Australia. http://www.aph.gov.au/Parliamentary_Business/Committees/Joint/Human_Rights_inquiries/FreedomspeechAustralia/Report (Accessed: 3 April 2017).

Paul, D. (1988) The selection of the "Survival of the Fittest", *Journal of the History of Biology*, 21(3), pp. 411-424. http://www.jstor.org/stable/4331067 (Accessed 2 March 2017).

Phillips, W. (2015) *This Is Why We Can't Have Nice Things: Mapping the Relationship between Online Trolling and Mainstream Culture.* Cambridge, Mass.: MIT Press.

Simon Wiesenthal Center (2017) *Simon Wiesenthal Center's 2017 Digital Terrorism & Hate Report Card: Social Media Giants Fail to Curb Online Extremism*, Simon Wiesenthal Center. http://www.wiesenthal.com/site/apps/nlnet/content2.aspx?c=lsKWLbPJLnF&b=4441467&ct=14988437 (Accessed 15 April 2017).

Solon, O. (2017) Tim Berners-Lee on the future of the web: "The system is failing", *The Guardian* (16 Nov), *TheGuardian.com*. https://www.theguardian.com/technology/2017/nov/15/tim-berners-lee-world-wide-web-net-neutrality (Accessed 16 March 2018).

Statt, N. (2017) YouTube is facing a full-scale advertising boycott over hate speech: The biggest brands continue to leave, *The Verge*. https://www.theverge.com/2017/3/24/15053990/google-youtube-advertising-boycott-hate-speech.

Stein, J. (2016) How trolls are ruining the Internet, *Time* (18 August/ 12 September) (Accessed 21 February 2017).

Suler, J. (2004) The online disinhibition effect, *Cyberpsychol Behav.*, 7(3), pp. 321-6, *Cyberpsychology and behaviour: the impact of the Internet, multimedia and virtual reality on behavior and society*. https://www.ncbi.nlm.nih.gov/pubmed/15257832 (Accessed 23 September 2016).

Tierney, S. (2007) *Multiculturalism and the Canadian constitution.* Vancouver: UBC Press.

UNESCO (1950/1954/1957/1969) *Four statements on the race question* Paris: Unesco. http://unesdoc.unesco.org/images/0012/001229/122962eo.pdf.

Zuckerberg, M. (2017) Building global community, *Facebook*. https://www.facebook.com/notes/mark-zuckerberg/building-global-community/10154544292806634/ (Accessed 2 May 2017).