CYBER RACISM, and COMMUNITY RESILIENCE

PROJECT REPORT AND RECOMMENDATIONS TO AUSTRALIAN HUMAN RIGHTS COMMISSION (AHRC) – CRaCR Team

Executive Summary

Preface

The report addresses priorities for the AHRC over the regulatory issues affecting the rapid spread of cyber racism. Research was undertaken by academics and associates in collaboration with the industry partners – the AHRC, VicHealth, FECCA and the Online Hate Prevention Institute.

Introduction

Online racism is an extension of racism more generally, compounded by the capacity of the Internet to allow the rapid and widespread distribution of racist hate speech. Even so, problems associated with defining racism and its regulation, remain.

- **Current Mechanisms for regulating cyber-racism in Australia**

The range of mechanisms operate at state and federal levels, in both criminal and civil jurisdictions, aimed at issues of racial vilification and intimidation, the use of telecommunications networks, and prevention of personal and social harm. The review of current mechanisms covers Federal and State/Territory Racial Vilification Laws in civil and criminal jurisdictions. It demonstrates that even where racism is covered by Federal law, differences among the states produce a very uncertain regulatory regime for individual citizens, varying significantly depending on the state in which the event is deemed to occur, the residency of the perpetrator and that of the target.

Criminal laws cover incitement to violence on the basis of race, and the use of Commonwealth licensed services to communicate harassing material. Harassment and public order offences are the focus of state criminal laws. Commonwealth media offences relate to the licensing of media organisations including online services, within the framework of traditional film and television censorship frameworks.

The report questions the efficacy of intermediary codes of conduct rules, which have been demonstrably inadequate in controlling racist hate speech online, except in the face of significant public and media campaigns.

The Commonwealth foreign affairs powers in relation to racism are constrained by the refusal of the Australian Parliament to authorise Australian accession to criminal sanctions on the basis of racial vilification.

- **Limitations of Existing Approaches**

Each of the current mechanisms has draw-backs. The civil provisions are onerous for the victims seeking cessation or redress, the criminal processes only operate to punish offenders, not to prevent the offence. A detailed assessment of each of the approaches identifies a significant gap. No single system exists that identifies and denounces race hate speech, and allows speedy and effective remediation.

- **Lessons from Existing Models for Regulating Comparable Harmful Online Conduct**

Other models considered include the E Safety response to cyber-bullying, where the offended party has to report the issue to the intermediary. If no action results the matter can be escalated. This is compared to the New Zealand situation, that empowers offended parties if required to take court action.

- **Closing the Regulatory Gap: Recommendations for a Civil Penalties Approach to Cyber-Racism**

Recommendations for action by the AHRC include that it move to having the government consider introducing a civil penalties approach, with a harm threshold that reflects community standards as evidenced in the public support for the key provisions of RDA Section 18C. Moreover a cyber bullying model, covering all ages, expanded to cyber racism and placing greater responsibility on the intermediaries to seek out and prevent the dissemination of hate speech would increase the role of the platforms in building a culture of civility online. Third party interventions should be allowed, so that civil society groups or the AHRC could commence action without the need for a complaint by an individual. Finally it is important to build a wider alliance between government, industry and civil society to develop strategies to resolve the issues raised in the report. The AHRC could take a leading role in promoting such an alliance, in conjunction with the other industry partners.

Preface

The Internet became "public" in about 1996. In the USA and elsewhere as it spread, organised racist groups hailed it as a game changer in the opportunities the public sphere offered for their projects of hate. Within months the United Nations groups concerned with racism were exploring how the world body might limit the damage the new technology was appearing to promote. That debate has continued for more than 20 years, with the extent and impact of the Internet now magnifying the corrosive effects of racism. In 2010 the Australian Human Rights Commission held a forum (its second in a decade) to explore cyber racism and determine what might be done in Australia to curtail its negative effect on community relations. The project from which this report is drawn, "Cyber Racism and Community Resilience (CRaCR)" grew out of the 2010 forum and the sense of impotence many participants felt in the face of the new technologies and their enhanced detrimental prospects, as social media erupted.

Our research (Jakubowicz et al., 2018) shows that the Internet's design was unfortunately suited to growing racism, for reasons to do with neo-liberal ideals of market freedom, democratic ideals of untrammelled communication, faltering and incompatible forms of regulation, competing self-concerned economic interests, fragmented community organisations, the tendency of users to feel disinhibited about their messaging compared with the offline world (Suler, 2004, Brown, 2017), and the nature of the technologies – from dispersed communication networks which are hard to control from a central point to social media algorithms that press towards serving advertising onto popular racist sites. The tensions between government, the commercial world and civil society over what could or should be done in relation to the infrastructure and rapidly evolving online services and software including social media platforms, lessened the capacity of more vulnerable social groups to resist their exposure to hate speech, intimidation and harassment. These targets and victims of online racism found it increasingly difficult to move past the growing impediments that both the Internet and negative discourses about cultural difference placed in the way of civility and cultural compatibility, thereby eroding the nascent possibilities offered by the emerging cyber democracy to build social capital in a multicultural society(Cohen-Almagor, 2011).

In 2012 the Australian Human Rights Commission agreed to join with VicHealth , the Federation of Ethnic Communities Councils of Australia, and with the Online Hate Prevention Institute as "industry" partners for the CRaCR research team – academics from three Victorian universities and three in NSW – to explore four dimensions of the increasing

social scourge of cyber racism. These elements – the encounters Internet users experienced with racism, the narratives that were emerging online about race and difference, the issues in regulation, and the potential strategies for building individual and community resilience – provide the formwork on which this report has been developed. The research was supported by an Australian Research Council Linkage Grant LP 120200115, funding and other inputs from the partners, and the participating universities. This report for the AHRC was prepared from the data developed under the Regulation stream of the project, extended through the Encounters stream, under the leadership of Prof Gail Mason, assisted by Ms Natalie Czapski, with input from Dr Andre Oboler.[1]

1

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).

Cohen-Almagor, R. (2011) 'Fighting Hate and Bigotry on the Internet', *Policy and Internet*, 3(3). http://www.psocommons.org/policyandInternet/vol3/iss3/art6 (Accessed 21 February 2017).

Jakubowicz, A., Dunn, K., Paradies, Y., Mason, G., Bliuc, A.-M., Bahfen, N., Atie, R., Connelly, K. and Oboler, A. (2018) *Cyber Racism and Community Resilience*. London: Palgrave Macmillan.

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).

CRaCR Team

Prof Andrew Jakubowicz
University of Technology Sydney

Dr Ana-Maria Bliuc
Western Sydney University

Prof Kevin Dunn
Western Sydney University

Dr Nasya Bahfen
Monash University

Prof Gail Mason
The Sydney Law School University of Sydney

Roslalie Atie
Western Sydney University

Prof Yin Paradies
Deakin University

Karen Connelly
University of Technology Sydney and Deakin University

Dr Andre Oboler
Online Hate Prevention Institute

Ms Natalie Czapski
The Sydney Law School University of Sydney

Provided for the use of the Australian Human Rights Commission.
Not for wider dissemination without consultation with authors.
Copyright rests with the Team and the AHRC
June 2017

Introduction

In this report we present the findings from the Regulation Stream of the ARC Cyber-Racism and Community Resilience project. We provide an overview of:

1. The Current Mechanisms for Regulating Cyber-Racism in Australia. This includes:
    1.1 Federal and State/Territory Vilification Law
    1.2 Criminal law
    1.3 The *Broadcasting Services Act 1992* (Cth) (BSA)
    1.4 Intermediary[2] Terms of Service and Codes of Conduct
    1.5 International Protocols and Standards
2. Limitations of the Current Approaches & Identification of the Gaps in Regulation
3. Lessons from Existing Models for Regulating Comparable Harmful Online Conduct
4. Closing the Regulatory Gap: Recommendations for a Civil Penalties Approach to Cyber-Racism

In many ways, the online expression of racism is simply an extension of racist conduct that occurs in the physical world, bringing with it the same regulatory challenges that lie at the core of all anti-racism policy and law. For instance, there is nothing about cyber-racism that avoids the definitional ambiguity surrounding racist speech or controversy about where the threshold of illegality should lie.

At the same time, however, the Internet provides unprecedented and novel opportunities for racism to flourish,[3] bringing new dimensions to the difficulty of regulation. The sheer volume of material and the speed of its dissemination to a wide audience means that isolated events and commentary can have global effects.[4] The opportunities for anonymity and the unmediated nature of the environment often circumvent regulatory mechanisms designed for traditional media outlets.

---

[2] Danielle Citron and Helen Norton use the term 'intermediary' to refer to private entities that host or index online conduct, such as Google, Facebook, YouTube and Twitter. We will adopt this term throughout this article. See Danielle Keats Citron and Helen Norton, 'Intermediaries and Hate Speech: Fostering Digital Citizenship For Our Information Age' (2011) 91(4) *Boston University Law Review* 1435, 1438-9. We also use the term 'content host'.

[3] Jesse Daniels, 'Race and racism in Internet Studies: A review and critique' (2012) 15(5) *New Media & Society* 695, 695; Lisa Nakamura and Peter A Chow-White (eds), *Race After the Internet* (Routledge, 2012) 17; Henry Jenkins, 'Cyberspace and Race: The color-blind Web: a techno-utopia, or a fantasy to assuage liberal guilt?' (2002) *MIT Technology Review* <http://www.technologyreview.com/article/401404/cyberspace-and-race/>.

[4] Brian McNair, 'When Terror Goes Viral It's Up to Us to Prevent Chaos', *The Conversation,* 27 July, 2016, <https://theconversation.com/when-terror-goes-viral-its-up-to-us-to-prevent-chaos-62687>

In Australia there is a spectrum of civil, criminal and voluntary systems that can be applied to regulate cyber-racism. At one end of the spectrum, telecommunications offences, and in rare instances, criminal vilification laws have been used to deal with cases of cyber-racism. At the other end of the spectrum, intermediary Terms of Service and Codes of Conduct provide a less formal and patchy avenue for redress. Undoubtedly, the lynchpin in Australia's approach to dealing with all forms of racist speech is the civil racial vilification model, which carries the 'practical regulatory burden' for both offline and online vilification.[5]

In the following report we analyse the relative merits of these systems to provide an effective and efficient process for handling complaints of racism in the online environment. Our goal is to assess whether there is a gap in the current regulatory responses to cyber-racism.

Despite the spectrum of regulatory mechanisms available in Australia, we conclude that there is a significant gap. There is no unified scheme that expressly denounces and remedies the harm of cyber racism by ensuring there is an efficient and accountable system for seeing harmful material removed, backed by a mechanism of enforcement.

Debates around freedom of speech and freedom from racism[6] highlight the wide diversity of views about how to define and respond to racist speech on the Internet. Such diversity may be best recognised through a multi-pronged approach that places greater regulatory responsibility on Internet companies that host content, while also offering aggrieved parties effective and enforceable avenues for confronting speech they find intolerable. Calls for such regulation are not new,[7] and looking to Australia's new cyber-bullying scheme,[8] may help us identify key elements for developing a comparable civil penalties approach for regulating cyber-racism.

---

[5] Katharine Gelber and Luke McNamara, 'The Effects of Civil Hate Speech Laws: Lessons from Australia' (2015) 49(3) *Law & Society Review* 631, 636.
[6] Peter Wertheim, 'Freedom and Social Cohesion: A Law that Protects both' (Paper presented at 40 Years of the Racial Discrimination Act 1975 (Cth) Conference, Sydney, 19-20 February 2015) 95 <https://www.humanrights.gov.au/our-work/race-discrimination/publications/perspectives-racial-discrimination-act-papers-40-years>; Attorney-General for Australia, 'Parliamentary Inquiry into Freedom of Speech' (Media Release, 8 November 2016) <https://www.attorneygeneral.gov.au/Mediareleases/Pages/2016/FourthQuarter/Parliamentary-inquiry-into-freedom-of-speech.aspx>.
[7] Andre Oboler, 'Time to Regulate Internet Hate with a New Approach?' (2010) 13(6) Internet Law Bulletin.
[8] *Enhancing Online Safety for Children Act 2015* (Cth).

1. The Current Mechanisms for Regulating Cyber-Racism in Australia

1.1 Federal and State/Territory Racial Vilification Laws

1.1.1 Civil Racial Vilification Laws

At the Federal level, s 18C of the *Racial Discrimination Act 1975* (Cth) (RDA)[9] makes it a civil wrong to do an act which is reasonably likely, in all the circumstances, to offend, insult, humiliate or intimidate a person or group on the basis of their race, colour, national or ethnic origin.[10] The act must be done 'other than in private'[11], which has been interpreted to include conduct occurring online, such as material published on a website that is not password protected.[12]

There is also racial vilification legislation in every state and territory, with the exception of the Northern Territory,[13] intended to operate concurrently with Commonwealth laws.[14] Most jurisdictions have both civil and criminal provisions; Tasmania has only a civil prohibition,[15] whilst Western Australia deals with racial vilification only through the criminal law.[16] The state and territory civil laws are largely based upon the NSW vilification legislation, which renders it unlawful for a person, by a public act, to incite hatred towards, serious contempt for, or severe ridicule of, a person or group of persons on the ground of the race of the person or members of the group.[17] Although the Victorian legislation is the only one to expressly include the use of the Internet or email to publish or transmit statements or material,[18] a 'public act' is broadly defined in the NSW legislation to include any form of communication.[19] It has been interpreted

[9] *Racial Discrimination Act 1975* (Cth) s 18C. Part IIA of the RDA, containing this racial vilification law, was implemented in order to give effect to Australia's obligations under the International Convention on the Elimination of All Forms of Racial Discrimination (ICERD).
[10] *Racial Discrimination Act 1975* (Cth) s 18C.
[11] *Racial Discrimination Act 1975* (Cth) s 18C(1).
[12] *Jones v Toben* (2002) 71 ALD 629, 646 [74], decision upheld in *Jones v Toben* [2003] FCAFC 137.
[13] In the Northern Territory, the *Anti-Discrimination Act 1992* (NT) makes it an offence to discriminate on the basis of race in education, employment, accommodation, goods, clubs, and insurance and superannuation, or on the ground that a person has a relative or associate of a particular race. However, there are no specific laws against racial vilification, although the Commonwealth vilification laws clearly apply in the Northern Territory as they do throughout Australia: *Racial Discrimination Act 1975* (Cth) s 18F; but see discussion in Neil Rees, Simon Rice and Dominique Allen, *Australian anti-discrimination law* (The Federation Press, 2nd ed, 2014) 617.
[14] *Racial Discrimination Act 1975* (Cth) s 18F; but see discussion in Rees, Rice and Allen, above n 12, 618-9.
[15] *Anti-Discrimination Act 1998* (Tas) ss 19-21.
[16] *Criminal Code 1913* (WA) ss 77-80D.
[17] Rees, Rice and Allen, above n 12, 670; *Discrimination Act 1991* (ACT) s 66; *Anti-Discrimination Act 1977* (NSW) s 20C; *Anti-Discrimination Act 1991* (Qld) s 124A; *Civil Liability Act 1936* (SA) s 73; *Anti-Discrimination Act 1998* (Tas) ss 19-21; *Racial and Religious Tolerance Act 2001* (Vic) s 7.
[18] *Racial and Religious Tolerance Act 2001* (Vic) s 7.
[19] *Anti-Discrimination Act 1977* (NSW) s 20B.

elsewhere to encompass the publication of material online.[20] In other words, racial vilification legislation generally applies to Internet content.

There are defences/exceptions to both Federal and state/territory civil vilification laws, including for certain types of material published reasonably and in good faith, such as academic publications, and fair and accurate reports on matters in the public interest.[21] The RDA does not operate extra-territorially.[22] However, it would appear, given the global nature of the Internet, that material which is uploaded or hosted overseas but can be viewed in Australia does fall within the bounds of the legislation.[23] A similar argument is likely to apply to state and territory vilification legislation.[24]

Under Federal legislation, the impact of the act is measured objectively from the perspective of a hypothetical reasonable person in the position of the applicant or the applicant's victim group, thereby applying community standards rather than the subjective views of the complainant.[25] It is sufficient to show that a particular sub-set of a racial group is reasonably likely to be affected by the conduct.[26] The conduct in question must cause 'profound and serious effects, not to be likened to mere slights'.[27] Conversely, the state/territory legislation considers the impact of the conduct on a *third party*, not the victim group. There is no need to prove that the respondent intended to incite or actually did incite anyone, provided that an ordinary member of the audience to whom it was directed would understand from the respondent's conduct that they were being incited towards hatred, serious contempt for, or severe ridicule of a person or persons, on the grounds of race.[28]

---

[20] In *Collier v Sunol* [2005] NSWADT 261, which involved homosexual vilification, the identical statutory definition of 'public act' was interpreted to include publication of material on the Internet.
[21] Australian Human Rights Commission (AHRC), 'Cyber Racism and Community Resilience Project, Working Paper: Civil and Criminal Racial Vilification Provisions' (Unpublished Working Paper, AHRC, July 2015) 8; *Racial Discrimination Act 1975* (Cth) s 18D; *Anti-Discrimination Act 1977* (NSW) s 20C(2); *Discrimination Act 1991* (ACT) s 66(2); *Anti-Discrimination Act 1991* (Qld) s 124A(2); *Civil Liability Act 1936* (SA) s 73(1); *Racial and Religious Tolerance Act 2001* (Vic) s 11.
[22] *Brannigan v Commonwealth* (2000) 110 FCR 566, 572-3.
[23] AHRC (2015), above n 20, 26; see *Dow Jones & Company Inc v Gutnick* (2002) 210 CLR 575, where material which was uploaded in the United States and downloadable by subscribers to a business news service in Victoria was held to have been published in Victoria for the purposes of defamation.
[24] The Victorian civil and criminal vilification provisions expressly apply to conduct occurring outside Victoria: *Racial and Religious Tolerance Act 2001* (Vic) s 7(1)(b); s 24(3)(b).
[25] AHRC, above n 20, 6; *Creek v Cairns Post Pty Ltd* (2001) 112 FCR 352, 356. See also *Eatock v Bolt* (2011) 197 FCR 261, 268, in which Bromberg J of the Federal Court concluded that the ordinary or reasonable hypothetical representative will have the characteristics that might be expected of a free and tolerant society.
[26] AHRC, above n 20, 6; see, eg, *McGlade v Lightfoot* (2002) 124 FCR 106, 117 [46]; *Eatock v Bolt* (2011) 197 FCR 261, 363 [452].
[27] *Creek v Cairns Post Pty Ltd* (2001) 112 FCR 352, 356 (Kiefel J).
[28] In *Catch the Fire Ministries Inc v Islamic Council of Victoria*, Ashley JA and Neave JA held that the effect of the conduct under Victorian legislation should be assessed with from the perspective of an ordinary member of the class of persons to whom the conduct was directed; Nettle JA preferred that it should be decided by reference to a *reasonable* member of that class. In *Sunol v Collier (No. 2)* (2012) 260 FLR 414, which considered homosexual vilification laws in NSW (equivalently worded to the NSW racial vilification laws), the Court of Appeal approved of the approach taken by the majority in Victoria. Cf. the approach taken to racial vilification legislation in *Veloskey v Karagiannakis* (2002) NSWADTAP 18 that one should consider the effect on the 'ordinary, reasonable reader'.

The harm threshold is therefore higher in the latter scenario, as the complainant must show that a third party, an ordinary, reasonable member of the general community rather than a hypothetical reasonable member of the victim group, could have been incited to feel hatred towards the victim group as a result of the respondent's conduct.[29] This is difficult to prove and less satisfactory for the victim, being divorced from their own personal reactions[30] or any assessment of the respondent's motive or intention in performing the act.[31] Incitement is also difficult to satisfy. Although it need not require evidence of causation, it does carry the connotation of 'inflame' or 'set alight' and is directed at conduct that is likely to generate strong and negative passions.[32] Accordingly, the ability of state/territory vilification laws to provide effective redress for those who feel aggrieved by speech they interpret as racist, including on the Internet, has been questioned.[33] There also continue to be gaps in the coverage afforded to religious vilification under both federal and state laws.[34] As we discuss further below, this is particularly problematic for Muslim Australians, who have been subject to an onslaught of Islamophobic behaviour in recent years, both online and off.[35]

In the majority of Australian jurisdictions, complaints of racial vilification are handled through a confidential process of conciliation wherever possible.[36] As an illustration, at the Federal level, the AHRC is responsible for investigating racial hatred complaints. Where the complaint cannot be resolved or is discontinued, the complainant is given documentation enabling them to instigate a complaint in Court;[37] equally, a civil case cannot be instigated unless a complaint has been made to the AHRC.[38]

---

[29] Rees, Rice and Allen, above n 12, 671; AHRC, above n 20, 21.
[30] AHRC, above n 20, 22.
[31] Rees, Rice and Allen, above n 12, 671; Australian Human Rights Commission, Submission to the Commonwealth Attorney General's Department, *Amendments to the Racial Discrimination Act 1975,* 28 April 2014, [97]-[99]; Victorian Equal Opportunity and Human Rights Commission, Submission to the Victorian Department of Justice, *Review of Identity Motivated Hate Crime*, May 2010, 16.
[32] AHRC, above n 20, 21.
[33] Ibid 22.
[34] Religion is only expressly protected in Queensland, Victoria and Tasmania. At the Federal level, s 18C of the RDA covers acts done on the basis of 'ethnic origin', which has been applied on numerous occasions to cover the vilification of Jewish people. There have been some comments in obiter to the effect that the law could cover Muslims, but this has not been tested: Eastman, above n 21, 143, citing *Jones v Scully* (2002) 120 FCR 243, 271-272 (Hely J); *Eatock v Bolt* (2011) 197 FCR 261, 333 [310].
[35] See, eg, Andre Oboler, 'Islamophobia on the Internet: The Growth of Online Hate Targeting Muslims' (Report IR13-7, Online Hate Prevention Institute, 10 December 2013); Mariam Veiszadeh, 'Muslim women scared to go outdoors in climate of hate', *The Sydney Morning Herald* (online), 11 October 2014, <http://www.smh.com.au/comment/muslim-women-scared-to-go-outdoors-in-climate-of-hate-20141009-113p5j>.
[36] Rees, Rice and Allen, above n 20, 627. The relevant Federal and state bodies include: Australian Human Rights Commission (AHRC); ACT Human Rights Commission; Anti-Discrimination Board of NSW; Anti-Discrimination Commission Queensland; Equal Opportunity Tasmania. South Australia employs a tort action civil law rather than a conciliation model as in other jurisdictions, and as of 2010, no complaints had ever been lodged under this law; see: Katharine Gelber and Luke McNamara, 'The Effects of Civil Hate Speech Laws: Lessons from Australia' (2015) 49(3) *Law & Society Review* 631, 641.
[37] Note that the AHRC does not provide any assistance with bringing or presenting Court proceedings. AHRC, *The Australian Human Rights Commission's Complaint Process: For complaints about sex, race, disability and age discrimination* (2016) <https://www.humanrights.gov.au/australian-human-rights-commission-s-complaint-process-complaints-about-sex-race-disability-and-age>.
[38] *Australian Human Rights Commission Act 1986* (Cth) s 46PO.

1.1.2. Criminal Racial Vilification Laws

Most jurisdictions also have a criminal offence of 'serious racial vilification',[39] adding to the civil requirements a further element: that the defendant must threaten physical harm to the person or property of the target person or group, or incite others to threaten harm of that kind.[40] Unlike the civil wrong, there are no statutory defences or exceptions. Again, the Victorian legislation expressly refers to the Internet and,[41] unlike the aforementioned jurisdictions, it extends to situations where the offender intentionally engages in conduct likely to incite serious contempt, revulsion, or ridicule without an aggravating threat of violence.[42] Western Australia, which differs markedly from other jurisdictions, takes an exclusively criminal approach. There are four offences concerning conduct that is intended to or likely to incite racial animosity or harass a racial group,[43] as well as corresponding strict liability offences with statutory defences.[44] Apart from one recent case in Queensland,[45] Western Australia is the only jurisdiction where there have been successful prosecutions for racially vilifying behaviour under vilification laws, including the conviction of a man who posted an anti-Semitic video on YouTube.[46]

There are no specific Commonwealth criminal offences concerned with racial vilification. However, it is a criminal offence to incite violence against a person or group of persons from a targeted group 'distinguished by race, religion, nationality, national or ethnic origin or political opinion.'[47] The applicability of this offence is narrow in relation to cyber-racism, the focus being on incitement of violence and not racist conduct per se.

1.2 Criminal Law

---

[39] *Discrimination Act 1991* (ACT) s 67; *Anti-Discrimination Act 1977* (NSW) s 20D; *Anti-Discrimination Act 1991* (Qld) s 131A; *Racial Vilification Act 1996* (SA) s 4; *Racial and Religious Tolerance Act 2001* (Vic) s 24; *Criminal Code* (WA) ss 77, 78, 79, 80, 80A, 80B, 80C, 80D.

[40] *Anti-Discrimination Act 1977* (NSW) s 20D; *Discrimination Act 1991* (ACT) s 67; *Anti-Discrimination Act 1991* (Qld) s 131A; *Racial Vilification Act 1996* (SA) s 4. NSW, Queensland and South Australia require consent to prosecute from the Attorney General (now delegated to the NSW Director of Public Prosecutions), a Crown Law Officer, and the Director of Public Prosecutions respectively.

[41] *Racial and Religious Tolerance Act 2001* (Vic) s 24.

[42] *Racial and Religious Tolerance Act 2001* (Vic) s 24(2).

[43] *Criminal Code* (WA) ss 78-80D. These includes offences involving possession of 'written or pictorial material', defined in in s 76 to mean 'any poster, graffiti, sign, placard, book, magazine, newspaper, leaflet, handbill, writing, inscription, picture, drawing or other visible representation.' This would presumably encompass materials on the Internet.

[44] *Criminal Code* (WA) ss 78, 80, 80B, 80D.

[45] Queensland teenager Abdel Kader Russell-Bouzmar pleaded guilty to a number of offences including serious racial vilification under s 131A of the *Anti-Discrimination Act 1991* (Qld) after an abusive tirade on a Brisbane train , and received a suspended sentence; see Kristina Harazim, 'Teen Abdel Kader Russell-Boumzar convicted over racially abusing guard on Brisbane train', *ABC News* (online), 14 September 2015, <http://www.abc.net.au/news/2015-09-14/teen-abdel-kader-russell-boumzar-convicted-brisbane-over-abuse/6775454>.

[46] *O'Connell v The State of Western Australia* [2012] WASCA 96 (4 May 2012); Standing Committee on Law and Justice, New South Wales Parliament Legislative Council, *Racial Vilification Law in New South Wales* (2013), [2.85].

[47] *Criminal Code* (Cth) ss 80.2A-80.2B.

1.2.1 Commonwealth Telecommunications Offences

Putting racial vilification offences to one side, the most obvious offence under which a person who puts racist material on the Internet might be charged is s 474.17(1) of the Commonwealth *Criminal Code*.[48] This makes it an offence to use a carriage service[49] in a way that reasonable persons would regard as being, in all the circumstances, menacing, harassing, or offensive. The offence has 'Category A' extended geographical jurisdiction,[50] meaning if the offender is an Australian citizen, they can be prosecuted even if the conduct occurred wholly outside of Australia.[51] The Explanatory Memorandum for the Amending Act[52] inserting the section makes it clear that the offence may be used to prosecute conduct that vilifies persons on the basis of race or religion.[53] Although there is no reported case law specifically concerning racially motivated conduct,[54] this section has been employed extensively to deal with harmful conduct online, with 308 successful prosecutions between its introduction in 2005 and 2014.[55]

Significantly, it has been directly and successfully applied in recent years to online conduct of a racially or religiously vilifying nature.[56] In 2014, a Western Australian man was charged with three counts under the section for a series of abusive tweets directed at an AFL player of Fijian heritage, in which he referred to the player as a 'black nigger', and referenced a desire to 'bash your black ass'.[57] The defendant ultimately pleaded guilty and received a conditional release order and $250.00 fine.[58] In 2016, a NSW chiropractor was convicted under the section for abusing Indigenous NT Senator Nova Peris, on

---

[48] A number of other provisions in Part 10.6 (Telecommunications Offences) may be applicable to cyber-racists, including s 474.15 (using a carriage service to make a threat to kill or cause serious harm) and s 474.16 (using a carriage service for a hoax threat).

[49] The Act provides that 'carriage service' has the same meaning as in the *Telecommunications Act*: that is, 'a service for carrying communications by means of guided and/or unguided electromagnetic energy.' The Internet, then, is clearly a 'carriage service.'

[50] This extended jurisdiction applies to all Part 10.6 offences, *Criminal Code* (Cth) s 475.2.

[51] *Criminal Code* (Cth) s 15.1.

[52] *Crimes Legislation Amendment (Telecommunications Offences and Other Measures) Act (No. 2) 2004* (Cth).

[53] Explanatory Memorandum, Crimes Legislation Amendment (Telecommunications Offences and Other Measures) Bill (No. 2) 2004 (Cth) 33.

[54] In *Starkey v Commonwealth Director of Public Prosecutions* [2013] QDC 124, the appellant successfully appealed his conviction under s 474.17 on the basis that the material in question was not sufficiently serious so as to engage the section. Inter alia, the appellant had sent emails containing Anti-Zionist and Anti-Semitic statements. See especially at [51] per Dorney DCJ.

[55] Explanatory Memorandum, Enhancing Online Safety for Children Bill 2014 (Cth), 52: see, eg, *Agostino v Cleaves* [2010] ACTSC 19 (3 March 2010) where the section was employed with respect to threats made by the accused via Facebook.

[56] This provision was also employed in the aftermath of the Cronulla Riots in 2005 to charge persons who sent text messages inciting the riots: see, eg, Kelly Burke and Ben Cubby, 'Police track text message senders', *Sydney Morning Herald* (online), 23 December 2005, <http://www.smh.com.au/news/national/police-track-text-message-senders/2005/12/22/1135032135717.html>.

[57] AAP, 'Man charged over racist tweets to Nic Naitanui', *PerthNow* (online), 21 March 2014, <http://www.perthnow.com.au/news/western-australia/man-charged-over-racist-tweets-to-nic-naitanui/story-fnhocxo3-1226861339231>.

[58] This was confirmed by records obtained from WA Magistrates Court.

Facebook, calling her a 'black c***' and demanding that she 'go back to the bush and suck on witchity [sic] grubs and yams'.[59]

While these unreported cases suggest that s 474.17 provides an avenue of redress for online vilification, the conduct in question must reach a high threshold of seriousness. In particular, the Crown must prove that a person (i) used a "carriage service" (ii) in a way that reasonable persons would regard as being, in all the circumstances, menacing, harassing or offensive.[60] In *Monis v The Queen*,[61] which dealt with the similarly worded s 474.12,[62] the High Court considered the requisite seriousness of conduct required to engage the section, noting that it protected against offensiveness 'at the higher end of the spectrum'.[63] In essence, this requires that the material be likely to (i) arouse significant anger, significant resentment, outrage, disgust or hatred in the mind of a reasonable person; (ii) cause a reasonable person to apprehend that the accused would cause the victim harm or injury; or (iii) cause a reasonable person to believe that the accused was troubling the victim/causing the victim apprehension by repeated attacks.[64] Accordingly, material that is still harmful may not reach the requisite level of seriousness required to engage the section unless it can be demonstrated that a reasonable person would respond in this way. In a recent Queensland District Court case, Dorney DCJ found that emails that were expressly Anti-Zionist or Anti-Semitic, including ones suggesting that certain groups or individuals should be 'shot by Humanity', could easily be accepted as offensive but did not meet the threshold for criminal sanction.[65]

In addition, there is a contention as to whether the 'reasonable person' referred to is one who merely has knowledge of the impugned material, or alternatively whether they are a reasonable person to whom the material was directed. An example of the former type of person would seem to be a reasonable white person who reads Internet material that is highly offensive to Sudanese migrants. An example of the latter type of person is the reasonable Sudanese migrant to Australia who reads the same material. French CJ noted this issue in *Monis*,[66] but the point was not raised in argument, and His Honour expressed no concluded view on the matter.

---

[59] ABC News, 'Man who abused Nova Peris on Facebook gets eight-month suspended sentence', *ABC News* (online), 5 July 2016, <http://www.abc.net.au/news/2016-07-05/man-who-abused-nova-peris-on-facebook-gets-suspended-sentence/7568912>.

[60] *Criminal Code* (Cth) s 474.3 sets out three matters to be included in a consideration of whether reasonable persons would regard the material as being 'offensive', those being (a) the standards of morality, decency and propriety generally accepted by reasonable adults, (b) the literary, artistic or educational merit (if any) of the material, and (c) the general character of the material (including whether it is of a medical, legal, or scientific character.

[61] *Monis v The Queen* (2013) 249 CLR 92, 123-4 [45].

[62] *Criminal Code* (Cth) s 474.12 is concerned with the use of 'a postal or similar service' rather than 'a carriage service', but is otherwise identical in wording to s 474.17.

[63] *Monis v The Queen* (2013) 249 CLR 92, 210 [336]. In *Brown v Commonwealth Director of Public Prosecutions* [2016] NSWCA 333, the NSW Court of Appeal found that there was no error in a primary judge applying the same construction of 'offensive' applied by the High Court in *Monis v The Queen* when construing s 474.17, given the identical wording of s 474.12 and s 474.17.

[64] *Monis v The Queen* (2011) 215 A Crim R 64, 77 [44] (Bathurst CJ); *Monis v The Queen* (2013) 249 CLR 92, 202-3 [310] (Crennan, Kiefel and Bell JJ).

[65] *Starkey v Commonwealth Director of Public Prosecutions* [2013] QDC 124, [51].

[66] *Monis v The Queen* (2013) 249 CLR 92, 123-4 [45].

There is some evidence that s 474.17 is an emerging regulatory 'frontier' for addressing some forms of cyber-racism. However, the practical utility of this promise is restricted to material that is 'likely to have a serious effect on the emotional well-being of the addressee' by arousing significant anger, resentment, outrage, disgust or hatred in the mind of a reasonable person.[67]

1.2.2 State and Territory Legislation

Other criminal provisions may also have application to racially motivated threats online. All jurisdictions have provisions concerning threats to kill,[68] as well as less serious threat offences, which vary by the level of threatened harm required to engage the offence.[69] The requirement of an imminent threat of harm largely rules out the applicability of assault offences to cyber-racism,[70] however, state based stalking and harassment offences could be used for the prosecution of cyber-harassment, including racial harassment.[71] In what was reportedly Australia's first prosecution of cyber-bullying, a man was convicted under the Victorian stalking legislation in 2010 over threatening text messages sent to a young person who eventually committed suicide.[72] A number of jurisdictions also have offences pertaining to the use of a computer system to publish or transmit objectionable material.[73] With some presently irrelevant differences between the legislation, each includes a prohibition relating to material that promotes crime or violence, or instructs in

---

[67] *Monis v The Queen* (2013) 249 CLR 92, 202-3 [310].

[68] *Crimes Act 1900* (ACT) s 30; *Crimes Act 1900* (NSW) s 31; *Criminal Code Act 1983* (NT) s 166; *Criminal Code Act 1899* (Qld) s 308; *Criminal Law Consolidation Act 1935* (SA) s 19(1); *Criminal Code* (Tas) s 162; *Crimes Act 1958* (Vic) s 20; *Criminal Code* (WA) s 338A.

[69] *Crimes Act 1900* (ACT) s 31; *Crimes Act 1900* (NSW) s 31; *Criminal Code Act 1983* (NT) s 200; *Criminal Code Act 1899* (Qld) s 359; *Criminal Law Consolidation Act 1935* (SA) s 19(2); *Criminal Code* (Tas) s 162; *Crimes Act 1958* (Vic) s 21; *Criminal Code* (WA) s 338B; see also *Crimes Act 1900* (NSW) s 199 (threatening to destroy or damage property). In some jurisdictions, threats to kill must be contained within a 'document' or put in 'writing', which would appear to encompass threats made on the Internet, for instance via an email or an online forum post, see, eg, *Criminal Code Act 1899* (Qld) s 1 (definition of 'document'); *Criminal Code* (Tas) s 1 (definition of 'writing').

[70] Gregor Urbas, 'Look who's stalking: cyberstalking, online vilification and child grooming offences in Australian legislation' (2007) 10(6) *Internet Law Bulletin* 62, 62; Simon Bronitt and Bernadette McSherry, *Principles of Criminal Law* (Thomson Reuters, 3rd ed, 2010) 563. In Tasmania, Queensland and Western Australia, words and images, whether online or otherwise, are insufficient evidence of a threat: *Criminal Code* (Tas) s 182(2); *Criminal Code Act 1899* (Qld) s 245; *Criminal Code* (WA) s 222; Des Butler, Silly Kift and Marilyn Campbell, 'Cyber Bullying and Schools in the Law: Is there an effective means of addressing the power imbalance?' (2009) 16(1) *Murdoch University Electronic Journal of Law* 84, 90.

[71] *Crimes Act 1900* (ACT) s 35; *Crimes Act 1900* (NSW) s 545B; *Crimes (Domestic and Personal Violence) Act 2007* (NSW) ss 7, 13; *Criminal Code 1983* (NT) s 189; *Criminal Code Act 1899* (Qld) ss 359A, 359B; *Criminal Law Consolidation Act 1935* (SA) s 19AA; *Criminal Code* (Tas) ss 192, 192A; *Crimes Act 1958* (Vic) s 21A; *Criminal Code* (WA) ss 338D; 338E. Legislation in all jurisdictions apart from Western Australia make some reference to the Internet or electronic or technologically assisted forms of communication.

[72] The accused pleaded guilty and received an 18-month community sentence, including 200 hours of unpaid community work: Selma Milovanovic, 'Man avoids jail in first cyber bullying case', *The Age* (online), 9 April 2010, <http://www.theage.com.au/victoria/man-avoids-jail-in-first-cyber-bullying-case-20100408-rv3v.html>.

[73] *Classification (Publications, Films and Computer Games) (Enforcement) Act 1995* (Vic) s 57; *Classification (Publications, Films and Computer Games) (Enforcement) Act 1996* (WA) ss 101-102; *Classification (Publications, Films and Computer Games) Act 1985* (NT) ss 77-78; *Classifications (Publications, Films and Computer Games) Act 1995* (SA) ss 75C, 75D(1).

matters of crime or violence,[74] as undoubtedly some racist material posted online would do.

Whilst every State and Territory has offensive language and offensive conduct provisions that are malleable enough to include racial vilification, they are rarely used to prosecute such conduct.[75] The legislation typically refers to conduct occurring 'in or near, or within hearing from, a public place or school' or similar,[76] and it is untested as to whether these references could be construed as extending beyond a physical locale to include the Internet as a publicly accessible space.[77]

1.3 The Broadcasting Services Act

Another possible avenue of recourse for cyber-racism is the online content scheme within Schedules 5 and 7 of the BSA. Regulated by the Australian Communications and Media Authority ('ACMA') until July 2015, responsibility for the scheme has now been assumed by the Office of the Children's eSafety Commissioner ('the Commissioner'), a separate, independent statutory office located within the ACMA.[78] The scheme imposes obligations upon Internet Service Providers (ISPs)[79] and content/hosting service providers in relation to certain harmful Internet content. When the provisions to regulate Internet content were first introduced in 2000,[80] it was clear that they did not deal specifically with racial hatred. Rather, the Act's Explanatory Memorandum made clear that its primary purpose was to protect children, and others, from Internet pornography.[81] However, it would seem that the scheme has some incidental application

---

[74] *Classification (Publications, Films and Computer Games) (Enforcement) Act 1995* (Vic) s 3; *Classification (Publications, Films and Computer Games) (Enforcement) Act 1996* (WA) s 99; *Classification (Publications, Films and Computer Games) Act 1985* (NT) s 75; *Classifications (Publications, Films and Computer Games) Act 1995* (SA) s 75A. In the South Australian legislation, objectionable material is defined to include Internet content consisting of a film or computer game that is classified RC, or would, if classified, be classified RC under the National Classification Code. This classification, and its lack of applicability to most 'everyday' instances of cyber-racism, is discussed in the context of the BSA in Section 3.3.

[75] David Brown et al, *Criminal laws: materials and commentary on criminal law and process of New South Wales* (The Federation Press, 2015) 541.

[76] *Summary Offences Act* 1988 (NSW) s 4 (Offensive conduct); s 4A (Offensive language); *Crimes Act 1900* (ACT) s 392 (Offensive behaviour); *Summary Offences Act* 1966 (Vic) s 17 (Obscene, indecent, threatening language and behaviour etc. in public); *Summary Offences Act* 1953 (SA) s 7 (Disorderly or offensive conduct or language); s 22 (Indecent language); *Criminal Code* 1913 (WA) s 74A (Disorderly behaviour in public); *Summary Offences Act* 1923 (NT) s 47 (Offensive conduct etc.); s 53 (Obscenity); *Police Offences Act* 1935 (Tas) s 12 (Prohibited language and behaviour).

[77] Interestingly the definition of a public place was held not to include the Internet in a Norwegian case, spurring their legislature to consider modifications to the Norwegian Penal Code: Nina Berglund, 'Lawmakers react to blogger's release', *News in English (Norway)* (online) 3 August 2012, <http://www.newsinenglish.no/2012/08/03/lawmakers-react-to-bloggers-release/>

[78] *Enhancing Online Safety for Children Act 2015* (Cth) s 15(1).

[79] This is defined in *Broadcasting Services Act 1992* (Cth) sch 5, cl 8(1) as those who supply, or propose to supply, an Internet carriage service to the public. The largest Australian examples include Telstra, Optus, TPG and Westnet, but there are many others: see Chris Connolly and David Vaile, 'Drowning in Codes of Conduct: An analysis of codes of conduct applying to online activity in Australia' (Final Report, Cyberspace Law and Policy Centre, The University of New South Wales, March 2012) 38.

[80] *Broadcasting Services Amendment (Online Services) Act 1999* (Cth).

[81] See Peter Coroneos, 'Internet Content Policy and Regulation in Australia' in Brian Fitzgerald et al (eds), *Copyright law, digital content and Internet in the Asia-Pacific* (Sydney University Press, 2008) 52.

to cyber-racism.

Schedule 5 is largely concerned with ISPs restricting access to content hosted overseas, in circumstances where the actual content hosts fall outside the Australian jurisdiction. In contrast, Schedule 7 deals with services that host or provide material online in or from Australia, collectively deemed as 'designated content/hosting service providers'.[82] Both ISPs and content/hosting service providers have obligations imposed upon them by two key regulatory codes registered with the ACMA (now the Commissioner).[83]

The scheme targets content that is 'prohibited' or 'potentially prohibited', meaning that it has been, or is substantially likely to be given an RC or X18+ rating by the classification board.[84] Under Schedule 7, the Commissioner must investigate a complaint into online content. If satisfied that the content meets this threshold, and is hosted in Australia, they may issue the provider with a 'take down notice', 'service-cessation notice' or 'link-deletion notice' as applicable.[85] Failure to comply with such a notice is an offence, as well as being a civil penalty provision.[86] Where the content meets this threshold, but is hosted overseas, Schedule 5 requires that the Commissioner notify ISPs who, according to the relevant Industry Code, are required to provide persons with access to filters to block content of this nature.[87]

Could racist content be covered by this scheme? The scheme is largely aimed at removing child pornography, but, as noted by the Australian Law Reform Commission,[88] the RC category is very broad. Under clauses 2, 3 and 4 of the *National Classification Code* (made in accordance with s 6 of the *Classification (Publications, Films and Computer Games)*

---

[82] Schedule 7 applies to 'hosting service providers', 'live content service providers', 'links service providers' and 'commercial service providers', collectively known as 'designated content/hosting service providers': see sch 7 cl 2 for definitions. Under cl 5, for the purposes of sch 7, 'a person does not provide a content service merely because a person supplies a carriage service that enables content to be delivered or accessed'. This provision would appear to exclude ISPs from obligations imposed on content and hosting services in the Schedule. It should be noted that there are other obligations on ISPs to do their best to prevent telecommunications networks and facilities from being used in, or in relation to the commission of offences against Australian law, and help must be provided to authorities as is reasonably necessary for the enforcement of the criminal law (which may include requests by authorities to block access to online services): *Telecommunications Act 1997* (Cth) s 313.

[83] Australian Law Reform Commission, *Classification – Content Regulation and Convergent Media: Final Report,* Report No 118 (2012) [2.29]; *Internet Association of Australia (IIA) Content Services Code of Practice* (Version 1.0, 2008); the *IIA Codes for Co-Regulation in the Areas of Internet and Mobile Content* (2005) ('the 2005 Code').

[84] *Broadcasting Services Act 1992* (Cth) sch 7 cls 20-21. Schedule 5 indicates that these terms are used in Schedule 5 as defined in Schedule 7. These classifications are in reference to the *National Classification Code*, made in accordance with s 6 of the *Classification (Publications, Films and Computer Games) Act 1995* (Cth).

[85] *Broadcasting Services Act 1992* (Cth) sch 7 pt 3 divs 3-5.

[86] *Broadcasting Services Act 1992* (Cth) sch 7 cls 106-7. Furthermore, if the Children's eSafety Commissioner (formerly the ACMA) is satisfied that a person is supplying a designated content/hosting service in contravention of the relevant provider rules, they may issue a written direction requiring the provider to take specified actions directed towards ensuring they do not contravene the rule in future. Failure to comply with a direction is both an offence and engages civil penalties: *Broadcasting Services Act 1992* (Cth) c 108.

[87] *Broadcasting Services Act 1992* (Cth) sch 5 cl 40; see also the 2005 Code, cls19.2, 19.3.

[88] Australian Law Reform Commission, *Classification – Content Regulation and Convergent Media: Final Report,* Report No 118 (2012) [2.29], 51; *Internet Association of Australia (IIA) Content Services Code of Practice* (Version 1.0, 2008); the *IIA Codes for Co-Regulation in the Areas of Internet and Mobile Content* (2005) ('the 2005 Code').

Act 1995 (Cth)), publications,[89] films and computer games, respectively, will be given an RC rating if they, relevantly:

> (a) ... deal with ... crime in such a way that they offend against the standards of morality, decency and propriety generally accepted by reasonable adults to the extent that they should not be classified; or
> ...
>
> (c) promote, incite or instruct in matters of crime or violence.

It seems conceivable that some online racist publications and films fit into either category (a) or (c) above (or both). Of the second of these two categories, the ALRC says:

> This means that material relating to drug use, shoplifting, graffiti or euthanasia could ... be classified RC.

Accordingly, this category is apparently broad enough to include, for example, a video (or written material) posted on the Internet that glorified acts of violence against Muslims. It might be that the same material would also fit into category (a) above. Furthermore, the Commissioner must inform the police if they believe content is of a sufficiently serious nature to warrant referral to a law enforcement agency.[90]

In sum, the Commissioner might be able to order that certain racist material on the Internet be taken down or (in the case of Internet content hosted overseas) filtered. But, if that is so, this is only true of material that has been, or is substantially likely to be, rated RC. It seems clear that not all racial material on the Internet would fall into this category, even if harmful. Overall, we must look elsewhere for regulation through which cyber-racist content can be dealt with effectively.

1.4 Intermediary Terms of Service and Codes of Conduct

Despite these Australian legal avenues, one of the most important paths of regulation for harmful content online are the terms of service and codes of conduct provided by intermediaries (private entities which host or link to online content). Online platforms typically have a set of terms that govern the behaviour of users that subscribe to their service, with stipulated mechanisms for reporting or dealing with harmful content. Many commentators champion the important regulatory role to be played by intermediaries, which are said to offer immediate, nuanced and flexible responses to 'hate speech' without the consequences associated with more 'heavy-handed' state action.[91]

---

[89] In s 5 of the *Classification (Publication, Films and Computer Games) Act* 1995 (Cth), 'publication' is defined as 'any written or pictorial matter' that is not a film, computer game or an advertisement for a publication, film or computer game.
[90] *Broadcasting Services Act 1992* (Cth) sch 5 cl 40(1)(a); sch 7 cl 69(1).
[91] Citron and Norton, above n 1, 1440-2.

There are numerous examples of intermediary terms of service that could address cyber-racist content. In the Australian context, individual ISPs have terms of service that implicitly, if not explicitly, encompass racist speech or the posting of racist content. For example, Optus prohibits the use of their service 'in any manner which improperly interferes with another person's use of our services or for illegal or unlawful purposes', including use of the service to 'defame, harass or abuse anyone'.[92] They reserve the right to block access to, remove, or refuse to post any content determined to be 'offensive, indecent, or otherwise inappropriate' regardless of whether it is lawful.[93] As this example shows, intermediaries often use language taken from legislation to articulate their terms of service but without fleshing out its defining features.

Of course, Australian ISPs and content hosts must follow Australian law, and therefore any cyber-racist material posted or accessed on such services is subject to the various legal mechanisms detailed above. The picture is more complex when we look at major content hosting platforms based overseas, and the way in which their terms of service operate and interact with our own legal system. After all, any provider code of conduct is voluntary, and there is no need for the platform to conform to the legislative requirements of any jurisdiction apart from the jurisdiction in which the service itself operates. Most of the world's largest social media platforms are based in the United States,[94] and are not prevented from restricting hate speech in the same way the First Amendment precludes government regulation.[95] Commentators argue that any reticence by these platforms to deal with harmful content can be read in light of the high value accorded to free speech in the United States, however repugnant or offensive.[96]

Facebook provides an illustrative example to examine the efficacy of platform terms of service for dealing with cyber-racist content. In addition to having over 13 million active users in Australia,[97] Facebook has been identified as a major site for the proliferation of cyber-racism.[98] All individuals and entities that make use of Facebook, and its various platforms and services, agree to the terms of service contained within Facebook's

---

[92] Optus, *Optus Fair Go Policy* (at 22 August 2016) s 4.
[93] Optus, *Optus Fair Go Policy* (at 22 August 2016) s 5.
[94] These include Facebook, Instagram (which is owned by Facebook and has over 500 million monthly active users), and YouTube, which is owned by Google.
[95] Citron and Norton, above n 1, 1439.
[96] Andre Oboler and Karen Connelly, 'Hate Speech: a Quality of Service Challenge' (Paper presented at IEEE Conference on e-Learning, e-Management and e-Services, Melbourne, Australia, 10-12 December 2014) 117, 118.
[97] Alex Heber, 'These incredible stats show exactly how huge Facebook is in Australia', *Business Insider Australia* (online) 8 April 2015, <http://www.businessinsider.com.au/these-incredible-stats-show-exactly-how-huge-facebook-is-in-australia-2015-4>.
[98] Facebook has been identified as one the main sites of cyber-racism in the preliminary results of the *Cyber-Racism and Community Resilience Project*, see Kevin Dunn, Yin Paradies and Rosalie Atie, 'Preliminary Results: Cyber Racism and Community Resilience: The Survey' (Unpublished results presented at Research Reporting Workshop, University of Technology Sydney, 28-9 May 2014), Slides 5, 10.

'Statement of Rights and Responsibilities'.[99] Under section 3 of the Statement ('Safety'), users undertake a commitment not to use the service to, amongst other specified terms:

- bully, intimidate, or harass any user
- post content that: is hate speech, threatening, or pornographic; incites violence; or contains nudity or graphic or gratuitous violence
- use[100] Facebook to do anything unlawful, misleading, malicious, or discriminatory.

Facebook's 'Community Standards'[101] are a further guide to the kind of behaviour and content that will not be tolerated. Relevantly for instances of cyber-racism, Facebook states that it will remove 'hate speech', which includes content that 'directly attacks people based on their: race, ethnicity, national origin, [or] religious affiliation.'[102]

Facebook's main tool for dealing with content that potentially violates its terms is via individual user reports. Facebook allows its users to flag material they find offensive and to report the content to Facebook for review.[103] Facebook indicates that they will review all reports of abusive and/or inappropriate content and remove content if they deem it to have violated the community guidelines. The 'Reporting Guide' indicates that most cases are reviewed within 72 hours, with more serious complaints being prioritised.[104]

Although Facebook's policies ostensibly prohibit racially abusive material and other kinds of harassing or offensive material, the social media platform has come under fire on numerous occasions for its failure to remove material that many people would deem offensive, or constituting 'hate speech'.[105] A pertinent Australian example is the 'Aboriginal Memes' page, which sprung up on Facebook in 2012 denigrating Indigenous Australians.[106] Facebook acknowledged that the page was 'incredibly distasteful' but initially refused to remove it on the grounds that it did not breach its terms of service.[107] The page was then briefly taken down, but re-emerged shortly afterwards with a

---

[99] Facebook, *Statement of Rights and Responsibilities* (30 January 2015) <https://www.facebook.com/legal/terms>.
[100] Facebook, *Statement of Rights and Responsibilities* (30 January 2015) s 18 <https://www.facebook.com/legal/terms> defines 'use' as 'use, run, copy, publicly perform or display, distribute, modify, translate, and create derivative works of.'
[101] Facebook, *Community Standards* (2016) <https://www.facebook.com/communitystandards>.
[102] Ibid.
[103] Facebook, *How to Report Things* (2014) <https://www.facebook.com/help/181495968648557>.
[104] Facebook, *What Happens After You Click 'Report'* (19 June 2012) <https://www.facebook.com/notes/432670926753695/>.
[105] See, eg, Andre Oboler, 'Islamophobia on the Internet: The Growth of Online Hate Targeting Muslims' (Report IR13-7, Online Hate Prevention Institute, 10 December 2013); Andre Oboler, 'Aboriginal Memes and Online Hate' (Report IR12-2, Online Hate Prevention Institute, 11 October 2012); Andre Oboler, 'Recognizing Hate Speech: Anti-Semitism on Facebook' (Report IR13-1, Online Hate Prevention Institute, 15 March 2013).
[106] The Race Discrimination Commissioner Helen Szoke at the time described the content as potentially being in breach of s 18C: Emma Skyes, 'Racist Facebook page deactivated after outcry', *ABC News* (online), 9 August 2012, <http://www.abc.net.au/local/stories/2012/08/08/3563446.htm>
[107] Online Hate Prevention Institute (OHPI), *PRESS RELEASE: Discussions with Facebook over Aboriginal Memes,* OHPI (15 August 2012) <http://ohpi.org.au/press-release-discussions-with-facebook-over-aboriginal-memes/>.

'[Controversial Humor]' tag. Further outcry saw an online petition for its removal gather over 20,000 signatures, and an investigation commenced by the ACMA.[108] It was finally removed by Facebook, apparently in response to this public pressure.[109] Even then, copycat pages continued to spring up, demanding ongoing intervention from Facebook.[110] In addition, social media platforms such as Facebook and YouTube often respond to content by blocking it within the jurisdiction in which it is potentially unlawful, rather than removing the material, meaning it can still be accessed overseas.[111]

These difficulties all point to the need for ongoing engagement between social media platforms and other intermediaries, governments and civil society, so as to better demarcate the socially acceptable grounds of behaviour online. Although major platforms appear compelled to remove material only when there is significant media or public backlash, they have demonstrated responsiveness to protracted government pressure to improve their moderation practices. In late 2015, the German government announced a landmark agreement with Facebook, Twitter and Google under which these platforms agreed to remove 'hate speech' from their platforms within 24 hours in Germany.[112] In early 2016, Facebook launched its European Online Civil Courage Initiative and pledged over €1 million to organisations and researchers seeking to understand and disrupt online extremism.[113] Any effort to better regulate cyber-racism must include and recognise the important role played by these intermediaries, in conjunction with jurisdictional legal frameworks.

1.5 International Protocols and Standards

Finally, it should be noted that there are a number of international protocols and standards that deal with harmful content online, including content of a racist or xenophobic nature. The core example of an intra-state regulatory framework is the *Additional Protocol to the Council of Europe Cybercrime Convention concerning acts of a racist and xenophobic nature committed through computer systems*. It requires state parties to criminalise 'making available' or distributing racist or xenophobic material through a computer

---

[108] Emma Skyes, 'Racist Facebook page deactivated after outcry', *ABC News* (online), 9 August 2012, <http://www.abc.net.au/local/stories/2012/08/08/3563446.htm>

[109] Ibid.

[110] Oboler, 'Aboriginal Memes and Online Hate', above n 104; Rod Chester, 'Facebook shuts vile Aboriginal memes page, despite earlier claiming it didn't constitute 'hate speech'', *News.com.au* (online), 27 January 2014, <http://www.news.com.au/technology/facebook-shuts-vile-aboriginal-memes-page-despite-earlier-claiming-it-didnt-constitute-hate-speech/story-e6frfrnr-1226811373505>.

[111] Oboler and Connelly, above n 95, 119. Government request statistics released by Facebook indicate that Facebook 'restricted accessed access in Australia to a number of pieces of content reported under local anti-discrimination laws' between July 2013 and December 2015: Government Requests Report, 'Australia', *Facebook* (2016) <https://govtrequests.facebook.com/country/Australia/2015-H1/>.

[112] Victor Lukerson, 'Facebook, Google Agree to Curb Hate Speech in Germany', *Time* (online) 15 December 2015, <http://time.com/4150296/facebook-google-hate-speech-germany/>. However, Lukerson states that it is unclear how this will affect comments made outside Germany that can still be read by German users, suggesting that an approach may be taken, similar to Google's approach to Europe's 'right to be forgotten' such that content can still be viewed on non-German versions of the platforms.

[113] Melissa Chan, 'Facebook Launches Initiative to Stop Extremist Posts in Europe', *Time* (online), 18 January 2016 <http://time.com/4184559/facebook-initiative-extremism/>.

system within their domestic laws.[114] Although Australia is party to the *Cybercrime Convention,* it declined to sign or ratify the Optional Protocol.[115] Outside the governmental sphere, the American-based Anti-Defamation League (ADL) has released 'Best Practice Guidelines for Challenging Hate', following consultation between Internet providers, civil society organisations, academics, and legal representatives.[116] These soft standards provide 'important guideposts' for both intermediaries and the Internet community at large.[117]

2. Limitations of the Existing Approaches & Identification of the Gaps in Regulation

2.1 The Racial Vilification Model: Definition, Confidentiality and Enforcement

As already noted, a significant area of uncertainty in the regulation of all forms of racial speech is the nuanced question of where to draw the threshold between speech that should be tolerated and speech that should be prohibited. The racial vilification model is the only system that attempts an explicit definition. Taking s 18C of the RDA as an example, material in the public domain is captured where it is offensive, insulting, humiliating or intimidating on the basis of race, colour, national or ethnic origin,[118] as measured objectively from the perspective of a hypothetical reasonable person in the position of the applicant or the applicant's victim group.[119] A strength of this threshold is that it does not prohibit generic offence or insult that confronts people with ideas or opinions with which they do not agree or which are mere slights to their feelings. Rather, it is restricted to comments that have 'profound and serious effects' that, arguably, impugn the dignity of people because of their race, colour or national or ethnic origin.[120] While some might see this as a narrow casting of the problem, this distinction attempts to balance free speech sensitivities with accountability for the harm of racial vilification.[121]

---

[114] *Optional Protocol to the Convention on Cybercrime, concerning the criminalisation of acts of a racist and xenophobic nature committed through computer systems,* opened for signature 28 January 2003, CETS No 189 (entered into force 1 March 2006) Art 3.
[115] Australian Human Rights Commission (AHRC), 'Human Rights in Cyberspace' (Background Paper, AHRC, September 2013) 24; Erik Bleich, *The Freedom to be Racist? How the United States and Europe Struggle to Preserve Freedom and Combat Racism* (Oxford University Press, 2011).
[116] Anti-Defamation League (ADL), *Press Release: ADL Releases "Best Practices" for Challenging Cyberhate,* ADL (23 September 2013) <http://www.adl.org/press-center/press-releases/discrimination-racism-bigotry/adl-releases-best-practices-challenging-cyberhate.html>.
[117] Anti-Defamation League (ADL), *Cyberhate Response: Best Practices for Responding to Cyberhate'* <http://www.adl.org/combating-hate/cyber-safety/best-practices/#.VIkCSGSUe5I>.
[118] *Racial Discrimination Act 1975* (Cth) s 18C.
[119] AHRC, above n 78, 6; *Creek v Cairns Post Pty Ltd* (2001) 112 FCR 352, 356. See also *Eatock v Bolt* (2011) 197 FCR 261, 268, in which Bromberg J of the Federal Court concluded that the ordinary or reasonable hypothetical representative will have the characteristics that might be expected of a free and tolerant society.
[120] *Creek v Cairns Post Pty Ltd* [2001] FCA 1007 at [16] per Kiefel J.
[121] Wertheim, above n 5, 100.

Despite these advantages, the grounds covered by s 18C remain limited. On the one hand, concepts such as race and ethnic origin have been construed broadly and are unlikely to be restricted to biological or racial markers.[122] For example, proof of ethnic origin can be established through evidence of factors such as membership of a population subgroup, common descent, national or cultural tradition, common language and migration status.[123] The application of the hate speech provisions to vilification based on religious beliefs remains 'unclear', presenting particular challenges for the inclusion of religions such as Islam that are not associated with any one ethnicity or race.[124] Eastman suggests that, if asked, the Federal Court 'may' find that a Sikh, Muslim or member of another minority religious community has an ethnic origin for the purposes of the RDA.[125]

Undoubtedly, the process of conciliation inherent to racial vilification laws can be advantageous, allowing for harmful conduct to be addressed quickly and informally, in a manner that educates the person who engaged in the conduct, without resort to the Court system. Victims of racist conduct often are not looking for the perpetrator to face heavy penalties, but simply a genuine apology acknowledging the harm and removal of the material.[126] In the context of cyber-racism, the company hosting the content may itself be the respondent. As demonstrated by the following case studies provided by the AHRC, intermediaries have shown some willingness to cooperate with the AHRC to remove racially vilifying material:[127]

- A complainant of Asian background reported a website that advocated violence against Asian people. The AHRC contacted the ISP to establish the identity of the

---

[122] The support of the Federal Court for Lord Fraser's approach in Mandla also suggests that 'ethnic group' may be given a contemporary meaning that is 'appreciably wider than the strictly racial or biological': *Mandla v Dowell-Lee* [1983] QB 1, 562. The exception to this broad approach to interpretation is the category of 'national origin', which has been interpreted as a status that is fixed at the time of birth: *Macabenta v Minister for Immigration & Multicultural Affairs* (1998) 90 FCR 202, 209-211.

[123] Kate Eastman, 'Mere definition? Blurred lines? The intersection of race, religion and the *Racial Discrimination Act 1975* (Cth)' (Paper presented at 40 Years of the Racial Discrimination Act 1975 (Cth) Conference, Sydney, 19-20 February 2015) 150 <https://www.humanrights.gov.au/our-work/race-discrimination/publications/perspectives-racial-discrimination-act-papers-40-years>. For example, ethnic origin under Australian law has been interpreted to include Jews: *Jones v Scully* (2002) 120 FCR 243, 272 [113]. Ethnic origin under English law has been interpreted to include Sikhs and Roma: *Singh v Rowntree Macintosh Ltd* [1979] ICR 504; *Commissioner for Racial Equality v Dutton* (1989) QB 783.

[124] Margaret Thornton and Trish Luker 'The Spectral Ground: Religious Belief Discrimination' (2009) 9 *Macquarie Law Journal*, 79-80. While Jews have been recognized as an ethnic group and covered by the *Racial Discrimination Act 1975* (Cth), Muslims have not.

[125] Eastman, above n 122, 150. Although the Explanatory Memorandum to the Racial Hatred Bill 1994 (Cth), 3, suggests that Muslims are to be included, it is only in obiter that the Federal Court has stated that s 18C may cover Muslims: Eastman, above n 122, 147. See *Jones v Scully* (2002) 120 FCR 243, 271- 72 [110]-[113] (Hely J) and *Eatock v Bolt* (2011) 197 FCR 261, [310] (Bromberg J). This is unlikely to be the case if religion is the only reason for the act, in the absence of any connection with ethnic origin: Eastman, above n 122, 14.

[126] Gelber and McNamara, above n 4, 647.

[127] AHRC, above n 20, 27. Conciliation data provided to us by the AHRC indicates that conciliations have occurred with respondent websites, ISPs and social media platforms.
See also AHRC, *Conciliation Register* (2016) <https://www.humanrights.gov.au/complaints/conciliation-register>.

website owner. Within a few days the website had been disabled by the ISP on account of it breaching their 'Acceptance Use Policy'.

- A complainant reported a user posting racially derogatory comments in a video posted on a file sharing website. When the website was contacted, the comments were removed and the offending user was suspended.

At the same time, because conciliation is a private and confidential process, it struggles to achieve the educational and 'standard setting' objectives which lie behind racial vilification legislation.[128] A small proportion of cases do proceed to a public hearing, creating important precedents that help set community expectations. However, only 2% of matters under civil vilification laws are resolved in this public manner.[129]

The online environment also raises particular challenges for the enforcement of racial vilification laws. Whilst these laws have been applied to material uploaded in a foreign jurisdiction, where that material can be viewed in Australia,[130] it is difficult to compel an overseas author to comply with Australian law. In addition, the author of racist material may be operating anonymously or under a pseudonym, requiring cooperation from or compulsion of host websites to identify them. As noted above, some intermediaries have informally complied with requests from the AHRC to remove offensive material.[131] Nonetheless, where that platform is hosted overseas, any formal order to disclose information can only be enforced with a corresponding order of the relevant counterpart jurisdiction, in accordance with their laws.[132]

There may also be difficulties in enforcing orders against third party host platforms directly. Although the failure of a host platform to remove cyber-racist material within a reasonable timeframe has the potential to contravene s 18C(1),[133] this depends on proof that this (in)action was connected with the race of the complainant. In *Silberberg v Builders Collective of Australia,* the respondent was not liable for failing to remove racist comments published on its site because this failure could not be causally connected to race.[134] Subsequently in *Clarke v Nationwide News*, the respondent *was* liable for racially vilifying user comments published underneath an article on their website, because they had solicited comments, put them through a moderated vetting process, and still allowed them to be published.[135] This raises the question of whether the necessary causal

---

[128] Luke McNamara, *Regulating Racism: Racial Vilification Laws in Australia* (Institute of Criminology, University of Sydney, 2002), 310.
[129] Gelber and McNamara, above n 4, 643, 646. In 2015-16, only 3% of complaints finalised by the AHRC were lodged in court: AHRC, *Racial Discrimination Complaints* (7 November 2016) < https://www.humanrights.gov.au/news/stories/racial-discrimination-complaints>.
[130] *Dow Jones & Company Inc v Gutnick* [2002] HCA 56.
[131] AHRC, above n 20.
[132] Ibid.
[133] *Silberberg v Builders Collective of Australia Inc* (2007) 164 FCR 475, 485 (Gyles J).
[134] Ibid, 486. In that case, the failure of an online forum administrator to remove racist material was just as easily explained by 'inattention or lack of diligence' than by any connection to the race or ethnic origin of the complainant. In relation to *Silverberg*, Huynor argues that even third party hosts who are aware of and refuse to remove offensive material may not be liable for their inaction under civil vilification laws: Jonathon Hunyor, 'Cyber-racism: Can the RDA Prevent It?' (2008) 35(7) *Brief* 30, 31.
[135] *Clarke v Nationwide News* [2012] FCA 307, [110].

connection could be established for major sites, such as Facebook, that do not pre-moderate. Nonetheless, the vicarious liability provision under s 18E of the RDA is an 'important weapon… [for holding] Internet service providers and social media platform providers to account for racist material they allow to remain published', by allowing an employer company to be made liable for the actions of an employee (e.g., a content moderator).[136]

Critically, there is evidence that this model places a heavy burden on the complainant to initiate and pursue proceedings.[137] Complaints cannot be brought by a third party or bystander, but must be brought by the victim or a representative of the victim group. Nor does any state authority have the ability to initiate a complaint or commence litigation.[138] Research by Gelber and McNamara into claims brought under civil vilification laws throughout Australia over a 20-year period found that successful complaints/litigation usually required an individual of extraordinary resolve to pursue the claim, backed by a well-resourced, respected community organisation.[139] Although conciliation may be mooted as a quick and efficient mechanism of dispute resolution, many complaints are terminated on account of procedural barriers and the lengthy time it takes in some jurisdictions to reach conciliation.[140] This is especially problematic in the online context, given the ease with which harmful material can proliferate.

One consequence of the difficulties in pursuing actions, especially in relation to complaints about online vilification, may be indicated by the reduction in online racism as a proportion of all claims. The AHRC recorded in its 2015/16 Annual Report, that online racism had declined to 8% of all race hatred complaints, and made up only 1% of all complaints filed under the *RDA*. This reduction may also reflect the increasingly responsive reactions by intermediaries to the initial complaints, avoiding action by the AHRC.

2.2 Criminal Law: Process, Dissemination and Individualisation

In some instances, the processes of the criminal law may be more effective in dealing with cyber-racism. One of the virtues of the criminal law is that the victim does not carry the enforcement burden and bystanders can play a role in bringing the matter to the

---

[136] Tim Soutphommasane, 'In defence of racial tolerance' (Speech delivered at the Australia Asia Education Engagement Symposium, Melbourne, Australia, 1 April 2014) <https://www.humanrights.gov.au/news/speeches/defence-racial-tolerance>. This still requires the relevant causal connection to be established under s 18C for the actions of the employee. It is distinct from an intermediary being held liable for 'inciting' or 'assisting' a third party to post racist content: the ancillary liability provision under s 17 of the RDA does not apply to the racial vilification laws: Hunyor, above n 200, 35.

[137] McNamara, above n 127, 310.

[138] Gelber and McNamara, above n 4, 637.

[139] Ibid 646.

[140] Ibid 643-4. This may not be the case in all jurisdictions. For example, in 2015-16, the average time it took the AHRC to finalise a complaint was 3.8 months: AHRC, *Racial Discrimination Complaints* (7 November 2016) <https://www.humanrights.gov.au/news/stories/racial-discrimination-complaints>.

attention of the police.[141] Complaints can be handled by the police, and, in some circumstances, dealt with more quickly than through protracted engagement in civil law conciliation schemes, or via reports to the individual host platform. For menacing, harassing or offensive conduct at the higher end of the spectrum of harm, the Commonwealth telecommunications offences that we discussed above appear to offer a satisfactory remedy for an increasing number of complainants.[142] At the less serious end of the spectrum, every State and Territory in Australia has offensive language and offensive conduct provisions that are malleable enough to include racial vilification. However, they are rarely used to prosecute such conduct.[143] The legislation typically refers to conduct occurring 'in or near, or within hearing from, a public place or school' or similar,[144] and it is untested as to whether these references could be construed as extending beyond a physical locale to include the Internet as a publicly accessible space.[145]

Putting the advantages of criminal law aside, a major limitation of the criminal law is that it contains no direct mechanism for stopping the dissemination of racist material or halting its re-production again and again. In addition, there are collective dimensions to the problem of online racism that are not well served by the traditional perpetrator/victim paradigm of criminal law. As McNamara argues, the most significant drawback of the criminalisation approach to the regulation of racial vilification lies in its 'individualising and marginalising effects' that remove racial vilification from 'its social context, and [deflect] attention from the harm suffered by members of the relevant group and the wider community'.[146] Unlike racial vilification law, which identifies and emphasises the specific impact of the conduct, most criminal offences do not explicitly name the harm or wrong of racism that we have identified above.[147]

---

[141] As demonstrated by the prosecution of a woman on a NSW train for offensive behaviour, which was filmed by a third party on a smart phone and posted online where it was widely condemned as racist: Lucy McNally, 1 August 2014, 'Karen Bailey gets good behaviour bond, avoid recorded conviction after racist train rant, http://www.abc.net.au/news/2014-07-31/karen-bailey-avoids-conviction-after-racist-train-rant/5638192, accessed 5 October 2016. Racially abusive comments on Facebook about Senator Nova Peris were also the subject of widespread media coverage and a swift public petition asking police to investigate the incident: Alina Tooley, *NSW police to investigate Chris Nelson for racial vilification of Nova Peris* (2016) Change.org <https://www.change.org/p/nsw-police-nsw-police-to-investigate-chris-nelson-for-racial-vilification-of-nova-peris>.

[142] *Monis v The Queen* (2013) 249 CLR 92, 202-3 [310].

[143] David Brown et al, *Criminal laws: materials and commentary on criminal law and process of New South Wales* (The Federation Press, 2015) 541.

[144] *Summary Offences Act* 1988 (NSW) s 4 (Offensive conduct); s 4A (Offensive language); *Crimes Act 1900* (ACT) s 392 (Offensive behaviour); *Summary Offences Act* 1966 (Vic) s 17 (Obscene, indecent, threatening language and behaviour etc. in public); *Summary Offences Act* 1953 (SA) s 7 (Disorderly or offensive conduct or language); s 22 (Indecent language); *Criminal Code* 1913 (WA) s 74A (Disorderly behaviour in public); *Summary Offences Act* 1923 (NT) s 47 (Offensive conduct etc.); s 53 (Obscenity); *Police Offences Act* 1935 (Tas) s 12 (Prohibited language and behaviour).

[145] Interestingly the definition of a public place was held not to include the Internet in a Norwegian case, spurring their legislature to consider modifications to the Norwegian Penal Code: Nina Berglund, 'Lawmakers react to blogger's release', *News in English (Norway)* (online) 3 August 2012, <http://www.newsinenglish.no/2012/08/03/lawmakers-react-to-bloggers-release/>

[146] See McNamara, above n 127, 247.

[147] The exceptions to this are the criminal racial vilification provisions, see above n 96. Although racial motive can be taken into account in some jurisdictions at sentencing, it is merely one aggravating factor amongst many: see, eg, *Crimes (Sentencing Procedure) Act 1999* (NSW) s 21A(2)(h).

2.3 Intermediary Terms of Service & Codes of Conduct

By way of contrast, Terms of Service and Codes of Conduct adopted by content hosts in the Internet industry offer self-regulatory arrangements that may be quicker and more flexible compared with the processes of both racial vilification and criminal law. Yet reliance on private entity terms of use raises its own difficulties. Being private entities, these services are not automatically beholden to the legal standards of non-host jurisdictions. Their responses, even when a complaint is upheld, are limited to removing content or suspending or terminating a user's account. Additionally, platforms may not have terms of service that adequately encompass cyber-racist behaviour to the standard expected under Australian law, or else may not adequately enforce those standards. Where such a platform fails to remove racist content, there may be little recourse for a victim of cyber-racism, especially where the apparent perpetrator uses a pseudonym or is located overseas.

There continue to be questions about the extent to which online content hosts could be directly liable for failing to remove racist material located on their platforms under current regulatory model in Australia. Developments overseas, in contrast, have seen an agreement reached between the European Commission and a range of platforms in which the platforms have committed to removing fifty percent of racist content within 24 hours.[148] It seems this agreement was developed as a compromise to avoid the European Commission or individual countries imposing greater liability on the platforms through law reform.[149]

2.4 Conclusion: A Gap

The Internet, by its very nature, presents significant regulatory challenges stemming from the quantity of activity, the technical capacity for amplification of messages including messages of racism, the ease of anonymity, its borderless nature and a culture where historically 'anything goes'.[150] The combined efforts of existing criminal, civil and industry schemes do provide options for redress, depending on variables such as the seriousness of the harm or whether the complainant is a member of the target group.

Nonetheless, even when the above systems are considered as a whole, it is apparent that there is a significant gap in the regulation of cyber-racism in Australia. There is no regime that expressly denounces and remedies the harm of racist speech by offering a speedy and efficient system for removing unacceptable online content, backed by a mechanism

[148] European Comission 'European Commission and IT Companies announce Code of Conduct on illegal online hate speech', http://europa.eu/rapid/press-release_IP-16-1937_en.htm
[149] OPHI, European Union Agreement with Social Media Platforms on Tackling Hate Speech, http://ohpi.org.au/european-union-agreement-with-social-media-platforms-on-hate-speech/
[150] Sarah Rohlfing, 'Hate on the Internet' in Nathan Hall et al (eds), *The Routledge International Handbook on Hate Crime* (Routledge, 2014) 293, 297.

of enforcement that engages hosts, perpetrators, bystanders and victims within a unified scheme. In the following sections we consider measures that might help address this gap.

3. Lessons from Existing Models for Regulating Comparable Harmful Online Conduct

Our project survey of Internet users shows that the most common way that they choose to respond to racism, whether as targets or witnesses, is 'within platform', that is, using Facebook as an example, by reporting the content and blocking or de-friending the author. [151] Of equal significance, is the finding that 80 per cent of survey respondents support laws against racial vilification, and 69 per cent support laws against religious vilification. [152] This suggests that Internet users want a spectrum of regulatory options that prioritise 'within platform' systems of complaint in the first instance, but backed by external legal mechanisms. It points to the need for a multi-pronged approach (one with several 'gears and levers') that provides alternative routes of redress for cyber-racism.

Though recommendations for greater control over online content often foreground additional criminalisation,[153] criminal law has not been the preferred vehicle of Australian legislatures for regulating racial vilification.[154] Yet, recently the Federal government has shown willingness to use *civil* mechanisms to regulate another form of harmful online material, namely, cyber-bullying directed towards Australian minors. [155] The New Zealand government has also enacted comparable but broader reforms through their *Harmful Digital Communications Act*.[156] Importantly, both schemes include mechanisms that engage with end-users (those who post harmful content), and the platforms which host harmful material.

3.1 The Australian Scheme for Regulating Cyber-Bullying

The federal government has recently introduced cyber-bullying laws, administered by the Children's e-Safety Commissioner.[157] The laws apply to seriously harassing, threatening,

---

[151] Cyber-Racism and Community Resilience (CRaCR) Research Group, Submission to the Commonwealth Attorney General's Department, *Amendments to the Racial Discrimination Act 1975,* 28 April 2014, 4. This is the same research cited earlier, conducted by the Encounters stream of the Cyber Racism and Community Resilience Project: above n 43-44.
[152] CRaCR Research Group, above n 150, 4.
[153] OHPI, above n 104, 63; Audrey Guichard, 'Hate Crime in Cyberspace: The Challenges of Substantive Criminal Law' (2009) 18(2) *Information and Communications Technology Law*, 201, 224. On the use of a carriage service for the dissemination of private sexual material or 'revenge porn' see Criminal Code Amendment (Private Sexual Material) Bill 2015 (Cth). See also Henry and Powell, above, n 209, 404.
[154] McNamara, above n 127, 204, 308.
[155] *Enhancing Online Safety for Children Act 2015* (Cth).
[156] *Harmful Digital Communications Act 2015* (NZ).
[157] *Enhancing Online Safety for Children Act 2015* (Cth). The government has proposed legislation that would rename the statutory office to that of the 'eSafety Commissioner', and expressly broaden the educative, research and advice-giving functions of the Commissioner to cover all Australians. This reflects the expanded role already being adopted by the Commissioner, and would not widen the scope of the cyber-bullying complaints system: Explanatory Memorandum, Enhancing Online Safety for Children Amendment Bill 2017 (Cth); Mitch Fifield, Minister for Communications, 'eSafety Office to Help All Australians Online' (Media Release, 9 February 2017)

intimidating or humiliating content targeted at Australian children.[158] Under the scheme, cyber-bullying material[159] must first be reported to the relevant 'social media service'.[160] For smaller services,[161] if the material is not taken down within 48 hours, the Commissioner can request that the material be removed, but there are no direct removal powers. For larger services - including Facebook, Google+, Instagram and YouTube[162] - the Commissioner can issue a notice requiring the service to remove the material within 48 hours. Failure to comply can result in a fine being issued, and if necessary, an injunction obtained in the Federal Court.[163] Persons who post the cyber-bullying material may be issued with a notice requiring them to remove the material, refrain from posting similar material, and/or apologise for posting the material.[164] If they do not comply, the Commissioner can issue a formal warning or obtain an injunction.[165]

Significantly, the Australian cyber-bullying regime places an initial obligation on persons to report harmful material to the relevant social media service – it is only when that material is not removed within a specified time that the Commissioner can either request or order the platform to remove it. In this way, the scheme recognises the processes of the social media service as the first port of call, but provides a backstop against inaction with the ability to enforce penalties against services and perpetrators.

<http://www.mitchfifield.com/Media/MediaReleases/tabid/70/articleType/ArticleView/articleId/1318/eSafety-office-to-help-all-Australians-online.aspx>.

[158] The legislation deals with material that is directed at a *particular* Australian child. It is therefore unable to account for cyber-racist material directed towards a certain racial group, of which the child is a member, as would be the case for much of the racist material being posted online.

[159] This is defined as material intended to effect a particular Australian child, where an ordinary reasonable person would conclude that such material would be likely to have an effect on that child of seriously harassing, threatening, intimidating, or humiliating them: *Enhancing Online Safety for Children Act 2015* (Cth) s 5(b).

[160] 'Social media service' is defined very widely under the legislation, and includes social networking platforms, blogging sites and apps, messaging apps which allow content to be included with messages, and video sharing sites and platforms; see *Enhancing Online Safety for Children Act 2015* (Cth) s 9; Office of the Children's eSafety Commissioner, *Information Guide: Cyberbullying Complaints Handing*, Commonwealth of Australia (Office of the Children's e-Safety Commissioner) (2015) 3 <https://www.esafety.gov.au/complaints-and-reporting/cyberbullying-complaints/complaint-resolution-process>.

[161] These smaller services are generally categorised as Tier 1 services. Any social media service may apply to be a Tier 1 social media service, which must be approved by the Commissioner provided that the service complies with basic online safety requirements as prescribed by the legislation, and provided the service is not a Tier 2 social media service; *Enhancing Online Safety for Children Act 2015* (Cth) ss 21, 23. The Office of the Children's e-Safety Commissioner website, as of December 2016, indicated that airG, Ask.fm, Flickr, Snapchat, Twitter, Yahoo!7 Answers and Yahoo!7 Groups are Tier 1 services: Office of the Children's eSafety Commissioner, *Social Media Partners: How our partners support the aims of the Office* (2016) <https://www.esafety.gov.au/social-media-regulation/social-media-partners>.

[162] Larger services are generally categorised as Tier 2 services. The Minister may declare a social media service to be a Tier 2 social media service on recommendation of the Commissioner. The Commissioner must not make a recommendation unless satisfied that the social media service is a 'large social media service', requiring an assessment of the number of accounts held by end-users resident in Australia and end-users who are Australian children, or on request of the social media service: *Enhancing Online Safety for Children Act 2015* (Cth) s 31; Office of the Children's e-Safety Commissioner, *Social Media Partners: How our partners support the aims of the Office* (2016) <https://www.esafety.gov.au/social-media-regulation/social-media-partners>.

[163] *Enhancing Online Safety for Children Act 2015* (Cth) ss 35, 36, 46-48.

[164] *Enhancing Online Safety for Children Act 2015* (Cth) s 42.

[165] *Enhancing Online Safety for Children Act 2015* (Cth) ss 43, 44, 48.

The scheme does this by allowing for court orders and potential fines on both large social media platforms[166] and on end-users who fail to remove cyber-bullying material.[167] Crucially, this new approach to cyber-bullying also integrates an educative function, with the Commissioner assuming a role in coordinating Commonwealth government efforts in online safety, education and research.[168] The legislation sets out basic online safety requirements for social media services.[169] In this way, content hosts are incentivised to improve their online safety and reporting practices.

3.2 The New Zealand Scheme for Regulating Harmful Digital Communications

The Australian legislation was inspired in part by the New Zealand legislation, which at that time was still being debated.[170] The New Zealand scheme employs an Approved Agency,[171] which is a designated non-government organisation, to investigate complaints about harm, defined as 'serious emotional distress', caused to individuals by digital communications.[172] Provided that an affected individual[173] has already brought a claim before the Approved Agency, they may bring a case in the District Court, which has the power to make orders against end-users and online content hosts.[174] Non-compliance with an order without reasonable excuse is a criminal offence.[175] Online content hosts may be insulated from proceedings by way of a safe harbour provision, which requires them to pass on a valid notice of a complaint about illegal content to the author of the content within 48 hours of receiving it, so as to give the author a chance to respond. If

---

[166] *Enhancing Online Safety for Children Act 2015* (Cth) ss 35, 36, 46-48.
[167] *Enhancing Online Safety for Children Act 2015* (Cth) ss 42, 43, 44, 48.
[168] *Enhancing Online Safety for Children Act 2015* (Cth) s 15.
[169] *Enhancing Online Safety for Children Act 2015* (Cth) s 21. If smaller platforms can demonstrate compliance with these and be approved by the Commissioner as 'Tier 1 Social Media Services', they are not subject to the punitive enforcement aspects of the legislation.
[170] Explanatory Memorandum, Enhancing Online Safety for Children Bill 2014 (Cth), 52.
[171] Netsafe, a New Zealand non-profit organisation promoting the safe use of online technologies, was officially appointed to fill the role in May 2016, and commenced work in November 2016. Amy Adams, 'Netsafe appointed to cyberbullying role' (Media Release, 31 May 2016) <https://www.beehive.govt.nz/release/netsafe-appointed-cyberbullying-role>; News and Communications, *Netsafe starts new role dealing with online harassment* (4 November 2016) New Zealand Law Society <https://www.lawsociety.org.nz/news-and-communications/latest-news/news/netsafe-starts-new-role-dealing-with-online-harassment>.
[172] *Harmful Digital Communications Act 2015* (NZ) ss 4, 7, 8.
[173] *Harmful Digital Communications Act 2015* (NZ) s 12(1). Claims may also be brought by a parent or guardian, a school on behalf of an affected student with consent, or the police per s 11.
[174] *Harmful Digital Communications Act 2015* (NZ) ss 18, 19. As per s 12(2), proceedings can only be brought if the District Court is satisfied that there has been a threatened serious breach, serious breach, or repeated breach of 1 or more communication principles, and the breach is likely to cause harm to an individual. Relevant communication principles for our purposes, as outlined in s 6, include:
- 2. Be threatening, intimidating, or menacing
- 8. Incite or encourage anyone to send a message to an individual for the purpose of causing harm to the individual
- 10. Denigrate an individual by reason of his or her colour, race, ethnic or national origins, religion, gender, sexual orientation, or disability.'
[175] *Harmful Digital Communications Act 2015* (NZ) s 21.

the author cannot be contacted or does not respond within 48 hours, the host must remove the content.[176]

3.3 Comparison of the Australian and New Zealand Schemes

Apart from the obvious difference in scope, the New Zealand scheme differs from the Australian cyber-bullying legislation insofar as the threshold of harm required to instigate proceedings is lower and less precisely defined. The legislation has been criticised for this reason. In contrast, the harm threshold set up by the Australian legislation - such that an ordinary reasonable person would conclude that material would be likely to have an effect of 'seriously harassing, threatening, intimidating or humiliating' a child[177] - is both less vague than the New Zealand standard and more in line with other Australian provisions regulating speech.

Both systems require a report to be made to a regulator outside of the systems provided by the social media platforms. The Australia system requires the user to first report the complaint to the hosting company using their internal systems. The New Zealand system does not require this. In both cases the regulator then contacts the social media company notifying them of a reported complaint. In the New Zealand system an onus is placed on the online content host to contact the user responsible for the content. This provides an additional educative impact on those who may have posted harmful content. The Australian regime by contrast educates those who are victims of abuse to report content to the platforms. The New Zealand legislation introduces a new broad criminal offence of causing harm by posting digital communication.[178] This is arguably unnecessary in Australia in light of existing Commonwealth telecommunications offences. One criticism of the New Zealand system is that platforms respond by simply taking down content reported to them by the regulator without any evaluation of the content. This shifts the cost of making a determination entirely away from the platform, which benefits financially from the system that also leads to these problems, and to the regulator, which is funded by the tax-payer or through public donations.[179]

The Australian cyber-bullying scheme provides a promising statutory model for exploring how best to confront the gap in protection from online racism. This potential was flagged in consultations prior to enactment of the legislation.[180] It presents a new and unparalleled opportunity to isolate a set of core elements that, with further detailed scrutiny, might prove translatable to the problem of cyber-racism. Next, we identify these elements and provide a nascent comment on their potential contribution to remedying the harm of cyber-racism.

---

[176] *Harmful Digital Communications Act 2015* (NZ) s 24.
[177] *Enhancing Online Safety for Children Act 2015* (Cth) s 5.
[178] *Harmful Digital Communications Act 2015* (NZ) s 22.
[179] Andre Oboler, "Online Safety Submission" (Online Hate Prevention Institute, 2014), https://www.communications.gov.au/sites/g/files/net301/f/submissions/Online_Hate_Prevention_Institute.pdf
[180] Ibid.

4. Closing the Regulatory Gap: Recommendations for a Civil Penalties Approach to Cyber-Racism

In this section we recommend seven elements to help close the regulatory gap around cyber-racism, illustrated by a brief hypothetical scenario on how these elements might come together to operate in practice. While we present these elements as a coherent model, this does not mean that a new bespoke civil enforcement regime is necessary. Individual elements have merit in their own terms and could be incorporated into existing schemes. We conclude this section with a discussion of the administration of a civil penalties approach to cyber-racism.

4.1 Recommendations

Recommendation 1: Articulation of a Harm Threshold that Reflects Community Standards

Any attempt to regulate cyber-racism should begin with a well-defined and appropriate threshold of harm for prohibited conduct. A strength of the Australian cyber-bullying legislation, over its New Zealand counterpart, is its articulation of a comparatively clear ambit. It applies to content that has a seriously harassing, threatening, intimidating or humiliating impact on a child, seemingly drawing together selected elements from the criminal law, including s 474.17(1) of the *Criminal Code* (Cth), and s 18C of the *RDA*.[181]

In our view, racial vilification laws offer the most logical starting point for determining an appropriate threshold of harm in any future civil penalty scheme prohibiting online racism. In particular, s 18C of the *RDA* sets a national standard [182] that has been in place for over 20 years. As we noted above, both Internet users[183] and the general community[184] have recently expressed support for the threshold of illegality set up by racial vilification laws. As McNamara and Gelber conclude, a 'very large majority of the public supports the idea that hate speech laws are an appropriate component of the framework within which public debate takes place in Australia.'[185] Nonetheless, no law

[181] While the scope of this legislation is confined to children as users of the Internet, such a limitation would be unduly restrictive in the context of cyber-racism which can inflict harm on persons of all ages.
[182] Tim Soutphommasane, 'A brave Act' (Paper presented at 40 Years of the Racial Discrimination Act 1975 (Cth) Conference, Sydney, 19-20 February 2015) 9 <https://www.humanrights.gov.au/our-work/race-discrimination/publications/perspectives-racial-discrimination-act-papers-40-years>.
[183] CRaCR Research Group, above n 150, 4.
[184] AHRC, *Overwhelming majority reject change to racial vilification law* (14 April 2014) <https://www.humanrights.gov.au/news/stories/overwhelming-majority-reject-change-racial-vilification-law>. While the Nielsen poll did not include the word 'intimidate', which is included in s 18C of the RDA, we might surmise that this is because intimidation is a breach of the criminal law and thus not a legal wrong that is peculiar to the *RDA*: see, eg, *Crimes Act 1900* (NSW) s 545B; *Crimes (Domestic and Personal Violence) Act 2007* (NSW) ss 7, 13.
[185] Luke McNamara and Katharine Gelber, 'The impact of section 18C and other civil anti-vilification laws in Australia', (Paper presented at 40 Years of the Racial Discrimination Act 1975 (Cth) Conference, Sydney, 19-20 February 2015) 167 <https://www.humanrights.gov.au/our-work/race-discrimination/publications/perspectives-racial-discrimination-act-papers-40-years>.

on its own will be able to fully reconcile strong differences of opinion over an appropriate threshold for intervention, which is why we call for a combination of legal and non-legal measures below.

To illustrate how this threshold element of a cyber-racism model might operate in practice, imagine, for example, that a new social media platform, 'AusBook', has been established. Whilst on the platform, a user comes across comments about a particular ethnic group, which employ highly derogatory language, brand the group as 'criminals and thugs', and insinuate that they should 'go back to where they came from'. To be captured by this model, and adopting the current s 18C threshold of harm, the comments would need to 'offend, insult, humiliate or intimidate' a person or group – in a way that has profound and serious effects that are more than mere slights - on the basis of their race, colour, national or ethnic origin. All of this would be assessed by a 'hypothetical' reasonable person in the position of the victim group. It would not be necessary to show incitement or subjective fault on the part of the person who made the comments.[186]

Although this threshold allows for a fairly broad interpretation of racist speech, one that is consistent with existing definitions of racial vilification, adopting the Commonwealth test does not resolve existing ambiguity and controversy at the definitional edges of racism. Much commentary that attracts the label 'racist' blurs the boundaries between race and other attributes or uses shorthand rhetoric to target one attribute while simultaneously referencing others. Slippage between categories of race, ethnicity and religion is a pertinent example. Notwithstanding claims that it should be 'beyond debate' that religious affiliation may be a marker of ethnic origin,[187] the 'racialisation of religious belief' has been criticised by Thornton and Luker for normalising some religions, such as Christianity, at the expense of others, particularly Islam. [188] While attempts to legislate for cyber-racism risk duplicating the definitional 'confusion and ambiguity'[189] that troubles all discrimination and vilification law, such uncertainty, and the controversy it attracts, is inherent to this field of law. It would be counterproductive to see it as an insurmountable obstacle to bespoke regulation of cyber-racism.

Recommendation 2: Utilisation of Existing Intermediary Reporting Mechanisms

---

[186] McNamara, above n 127, 207.

[187] Eastman, above n 122, 150. It must be noted, however, that despite comparable parliamentary intention in NSW, Muslims have not been included in the term 'ethno-religious origin' which was inserted into the *Anti-Discrimination Act 1977* (NSW) to 'clarify that ethno-religious groups such as Jews, Muslims and Sikhs have access to racial vilification and discrimination provisions in the Act': see New South Wales, *Parliamentary Debates,* Legislative Assembly, 4 May 1994, 1827 (John Hannaford). Nonetheless, the decision in *Khan v Commissioner, Department of Corrective Services* [2002] NSWADT 131 has meant that Muslims generally are not protected by the ADA in NSW: Eastman, above n 19, 149. Vilification based on religious belief is more explicitly prohibited in Victoria, Queensland and Tasmania: *Racial and Religious Tolerance Act 2001* (Vic) ss 8, 25; *Anti-Discrimination Act 1991* (Qld) ss 124A, 131A; *Anti-Discrimination Act 1998* (Tas) s 19.

[188] Thornton and Luker, above n 123, 74, 91.

[189] Ibid 72.

Attempts to tighten the regulation of racial comments online cannot ignore the key role played by hosts in dealing with harmful content, particularly as they are the preferred avenue of complaint for many users. The advantage of the cyber-bullying model is the requirement that people first make reports about harmful content directly to online content hosts, relying upon their existing terms of service and reporting mechanisms. For example, under our hypothetical AusBook scenario, any user concerned about the derogatory nature of comments towards the ethnic group in question would be required to report the content to the AusBook platform, through their content violation reporting mechanisms. It is only where this proves ineffectual or inapplicable that further intervention is possible under this scheme. This has the effect of ensuring the companies who provide a platform and financially profit from doing so also pick up the bulk of the financial cost of regulating their platform.

The shortcoming of this element of the scheme is that it places responsibility on the individual to make an initial complaint, and content hosts to respond satisfactorily. Yet, the sheer volume of Internet content makes it impossible for any private or public agency to regularly monitor it, effectively leaving the initial identification of unacceptable material in the hands of individual users.

Recommendation 3: Pressure on Content Hosts to More Effectively Police Online Conduct, Including Liability for Failure to Respond

Enacting firm legislation of this sort puts pressure on content hosts to improve their mechanisms for dealing with racist conduct and enforce existing codes more effectively, so as to avoid civil penalties.[190] The potential of this kind of scheme for cyber-racism is evidenced in the apparent success of the cyber-bullying legislation in its first six months of operation. Under that legislation, larger Tier 2 social media platforms risk an $18,000 fine/day if they fail to comply with a takedown notice within 48 hours, but hearteningly, most have responded in less than a day, meaning there is no need to fall back on civil penalties.[191]

Under our hypothetical example, this means that if AusBook, having received a complaint, decides that the material does not violate their terms of service because, for instance, the comments do not amount to a direct threat to any individual person(s), they may choose not to remove the material. In such circumstances, the user can report the content to the relevant statutory body, which then makes a determination as to whether the comments breach the requisite threshold of harm. If so, AusBook would be issued with a notice to take down the content within a certain period, or risk an injunction and/or a fine for non-compliance.

---

[190] Andre Oboler, 'Time to Regulate Internet Hate with a New Approach?' (2010) 13(6) Internet Law Bulletin.

[191] Office of the Children's eSafety Commissioner, 'Annual Report 2015-16' (Report, 2 September 2016) 122-4; Sunanda Creagh, 'Response from a spokesperson for Mitch Fifield', *The Conversation*, 29 August 2016 <http://theconversation.com/full-response-from-a-spokesperson-for-mitch-fifield-64439>.

By placing responsibility on content hosts to respond to complaints, including the removal of material, the advantage of this element is that it by-passes the need for an identifiable perpetrator. This gives it an edge over the racial vilification system in circumstances where the authors of material can be difficult to identify or out of jurisdictional reach. Importantly, the incorporation of a system for enforcing the prompt removal of material helps minimise the harm of racial vilification online.

Smaller platforms can be exempt from enforcement provisions if they demonstrate their terms of service clearly address the harmful behaviour. In the cyber-bullying legislation, for example, smaller platforms are given 'Tier 1' status, reflecting the commitment of the scheme to work in 'partnership' with social media services.[192] While this has the potential to create disparity in the enforcement mechanism, Tier 1 status can be revoked if a service fails to comply with basic online safety requirements, including if there is a repeated failure to respond to requests to remove material. This thereby subjects the service to the enforcement mechanism.[193] However, the efficiency of this process for unresponsive hosts is untested.

A further shortcoming of the cyber-bullying enforcement mechanism is that the maximum $18,000 daily penalty is a very small 'stick' given the high profit margins of major online platforms.[194] A higher penalty would certainly not be unreasonable to deal more effectively with harmful online content.

Recommendation 4: Allowance for Third Party Intervention

Like the cyber-bullying system, any civil penalty scheme for cyber-racism would need to be drafted in such a way that complaints could be brought by third parties / bystanders, or even by the State, where cyber-racist content is identified. In other words, our hypothetical AusBook complainant would not need to be a member of the vilified group in order for their complaint to proceed. This would distinguish the regime from existing conciliation procedures and civil vilification laws, which require an affected victim or victim group to initiate the claim. A crucial component of anti-racism strategies is giving bystanders the ability to call out and respond to racism, and in doing so, influence the mentalities around its perpetration.[195] Involving non-victim parties in any regulatory scheme would help build community capacity to identify and respond to racist behaviour and take the pressure off those who are its targets.

---

[192] Office of the Children's eSafety Commissioner, 'About Tier 1 of the scheme' (2016) <https://www.esafety.gov.au/social-media-regulation/about-tier-1-of-the-scheme>
[193] Ibid.
[194] The most significant example of this is Facebook, which reported profits in excess of $1 billion US in the final quarter of 2015: Deepa Seetharaman, 'Facebook Profit Tops $1 Billion', *Wall Street Journal* (online), 27 January 2016 <http://www.wsj.com/articles/facebook-profit-tops-1-billion-1453929139>
[195] Rivkah Nissim, 'Building Resilience in the Face of Racism: Options for Anti-Racism Strategies' (2014) *Sydney Social Justice Network: Australian Policy Online* <http://apo.org.au/resource/building-resilience-face-racism-options-anti-racism-strategies> 4.

In relation to civil vilification laws, McNamara argues that it is not the emphasis on conciliation that makes this scheme effective, but 'the relative ease with which proceedings to invoke legislative standards can be commenced and conducted', combined with the relatively broad scope of the legislation. [196] By the same token, the cyber-bullying scheme provides a simple online complaint process that has the potential to offer a less protracted or onerous path to resolution, for example by allowing the Commissioner to act on a complaint, and to act quickly, if material is not removed.[197] By minimising the burden on individual victims this kind of 'collective response'[198] goes some way towards recognising that racial vilification constitutes a public, not just an individual, wrong.

Recommendation 5: Penalties for Perpetrators of Cyber-Racism

Where the perpetrator is identifiable, the State can also intervene with the threat of civil penalties to enforce any orders made. For example, if the end-user who posted the derogatory material in our AusBook hypothetical could be identified, the responsible statutory body could also issue them with a takedown notice. One option would be to issue a fine that is heavily 'discounted' if both payment and the removal of content occur within a short period of time from the issuing of the notice.

This improves upon the conciliation model, for which enforcement against perpetrators remains a barrier to efficacy. In some cases, having a scheme that acts as an alternative to conciliation may be preferable, particularly where the complainant might not wish to confront the perpetrator, or where the perpetrator is not willing to engage in that process (or is otherwise unable to be found). Perpetrators who recognise their mistake may also prefer removing the content and paying a small fine than engaging in protracted discussion. Having a legal threat overhead, and an incentive for positive action, is likely to catalyse end-user compliance with orders to take down material without having to resort to a Court settlement to enforce the orders made.

Recommendation 6: Mechanisms to Educate Internet Users

---

[196] McNamara, above n 127, 310.

[197] Office of the Children's eSafety Commissioner, 'Complaints and Reporting' (2016) <https://www.esafety.gov.au/complaints-and-reporting>. Taking nearly five years for the complaint to be resolved, *Trad v Jones & Anor (No. 3)* (2009) NSWADT 318 is an example of a protracted complaint under the civil human rights system of conciliation; on subsequent legal proceedings and the question of costs see Gelber and McNamara (2015), above n 4, 648. Like the civil rights conciliation model, this approach also has a comparative advantage over the criminal jurisdiction, which requires people to lodge their initial complaint with the police, who are themselves the subject of distrust amongst some ethnic minorities and Indigenous Australians for perceived racial bias: Diane Sivasubramaniam and Jane Goodman-Delahunty, 'Ethnicity and trust: perceptions of police bias' (2008) 10(4) *International Journal of Police Science & Management* 388, 388-9.

[198] Katharine Gelber and Luke McNamara, 'Private litigation to address a public wrong: A study of Australia's regulatory response to "hate speech"' (2014) 33(3) *Civil Justice Quarterly* 307, 333-4.

Despite the benefits of the confidential conciliation process used in civil vilification laws, [199] research suggests that the public continue to lack knowledge about the rules that govern both racial vilification[200] and harmful online speech.[201] This reinforces the need for an educational function to be built into any regulatory scheme for cyber racism.

By encouraging access to online reporting mechanisms, the cyber-bullying approach can help equip users, including many who are young people, [202] with an understanding of appropriate standards against which to identify and respond to racial commentary, whether as the target or the witness of such speech.[203] The advantages can already be seen in the voluntary reporting system FightAgainstHate.com and the publications of the Online Hate Prevention Institute which explain racist content and encourage its reporting. Orders made under the proposed scheme would play a similar role, being available in the public domain to play an educative role about inappropriate online behaviour.[204] If these elements were built into any future cyber-racism scheme they could be used to foster community standards around racist speech, feed into broader educative initiatives by existing agencies and, as civil law remedies, potentially 'target a wider range of expressive conduct than a purely criminal model would permit'.[205] AusBook, for example, could be advised to update their terms of service to encompass vilifying material that did not amount to direct threats. This would also promote their conformity with other Australian civil and criminal standards around harmful content.[206] While platforms have traditionally been reluctant to adjust terms of service to comply with local law, since efforts by Germany in 2015 and 2016, they are increasingly doing so where governments insist it is required.

Recommendation 7: Enhancement of the Ability to Record and Monitor Online Behaviour

---

[199] Gelber and McNamara, above n 4, 643.
[200] AHRC, above n 114, 15; Gelber and McNamara, above n 4, 643.
[201] New Zealand Law Commission, above n 50, [3.43]-[3.52].
[202] A survey of social media use in Australia found that over 95% of the population aged between 18 and 29 accessed the Internet daily, with 79% of that age bracket using social networking sites at least once per day. ACMCA research into Australian teenagers using the Internet found that 72% accessed the Internet more than once a day, with Facebook and YouTube amongst the most accessed platforms. See Sensis, 'Sensis Social Media Report: May 2015: How Australian people and businesses are using social media (Report, Sensis and the Digital Industry Association of Australia, May 2015) 11, 14; Erin Raco, 'Research snapshots: Aussie teens online', *Australian Communications and Media Authority* (01 July 2014) <http://www.acma.gov.au/theACMA/engage-blogs/engage-blogs/Research-snapshots/Aussie-teens-online>.
[203] Nissim, above n 194, 3.
[204] See, for example, the educative role played by the Office of the Children's E-Safety Commissioner, as well as the Australian Cybercrime Online Reporting Network (ACORN), which was launched in November 2014 to offer a simple and streamlined way to report cybercrime and ensure reports are referred to the correct agency (http://www.acorn.gov.au/): Paul Osborne, 'New online tool ACORN allows Australians to report cybercrime in real time', *The Sydney Morning Herald* (online), 26 November 2014, <http://www.smh.com.au/digital-life/consumer-security/new-online-tool-acorn-allows-australians-to-report-cybercrime-in-real-time-20141125-11u0v1.html>.
[205] Gelber and McNamara, above n 4, 649.
[206] See, eg, *Racial Discrimination Act 1975* (Cth) s 18C; *Criminal Code* (Cth) s 474.17.

A final element of the cyber-bullying scheme that would be helpful in the context of cyber-racism is that it provides a channel to record and monitor activity as it is reported, adding to existing efforts in this area.[207] For example, in the hypothetical case of Ausbook, a published outcome report would alert other fledging platforms to this issue, whilst adding to data about the types of vilifying material being posted online, the 'usual' targets, and the common compliance-gaps faced by content hosts. This information has the potential to enhance community understanding about appropriate online standards. Empirical work of this nature has been recognised as critical and has been a high priority since 2009.[208] Despite this, the first, and so far only, report of this kind across multiple platforms was released in 2016 and examined both the prevalence of antisemitic content and the time it took Facebook, YouTube and Twitter to remove this content through their self-regulation mechanisms.[209] This sort of approach is needed on an ongoing basis across all forms of cyber-racism if self-regulation is going to improve in both effectiveness and efficiency. The AHRC should consider developing a collaboration with other agencies, industry groups and non-government organisations to effectively support the Online Hate Prevention Institute to ensure independent and effective monitoring of value to the three sectors.

4.2 The Administration of a Civil Penalties Scheme

This leaves the question of which existing agency would be best placed to administer a civil penalties scheme – or selected elements of such a scheme - for cyber-racism. The reporting, enforcement and penalty mechanisms built into the model under discussion here place it outside the current authority of the AHRC. Yet the proposed harm threshold and educational principles are a perfect fit with the mission of the AHRC to promote and protect human rights, as well as with its statutory responsibility for dispute resolution, public education and policy development.[210] One option would be to bolster the existing powers and legislative obligations of the AHRC to administer aspects of a complaints and compliance system along the lines we have proposed here.

Tasked with ensuring that media and communications legislation and codes of conduct operate 'effectively and efficiently, and in the public interest',[211] ACMA may also be an appropriate agency. It has significant authority over the development of codes of conduct, complaint processes and commercial broadcasting licenses for radio and

---

[207] See, eg, Online Hate Prevention Institute (OHPI), *Fight Against Hate* (2015) <https://fightagainsthate.com/>.
[208] Andre Oboler and David Matas, *Online Antisemitism: A systematic review of the problem, the response and the need for change* (Israeli Ministry of Foreign Affairs, 2013) 32. < http://mfa.gov.il/MFA/AboutTheMinistry/Conferences-Seminars/GFCA2013/Documents/OnlineAntisemitism.pdf>; Andre Oboler, "Measuring the Hate: The State of Antisemitism in Social Media" (Online Hate Prevention Institute, 2016). Online: http://ohpi.org.au/measuring-antisemitism/ 2
[209] Andre Oboler (2016), above n 207.
[210] Australian Human Rights Commission, 'Corporate Plan 2015-2016', 6.
[211] ACMA, 'Introduction to the ACMA' (5 September 2016) <http://www.acma.gov.au/theACMA/About/Corporate/Authority/introduction-to-the-acma>.

television.[212] In addition, it is already empowered to undertake enforcement action, including applications to the Federal Court for certain orders and civil penalties.[213] There is also a range of criminal, civil and administrative penalties within the *BSA*. The express expansion of these powers to online content including racist speech, and perhaps other forms of prejudicial content, would seem a natural fit.

The Office of the Children's eSafety Commissioner might also be an appropriate option. The Commissioner is concerned with online safety in a number of domains outside the cyber-bullying area, including the administration of the online content scheme under the BSA, and the promotion of women's safety online.[214] The proposed legislative changes also show an intention to formalise the expansion of the office to promote online safety for all Australians, not just primarily children.[215] The elements proposed above thus nicely compliment the functions of the Commissioner.

4.3 Conclusion

In sum, this 'broad brush stroke' exploration of the extent to which key elements of the new cyber-bullying scheme might meet comparable regulatory needs in the context of cyber-racism is not intended to be a forensic analysis of a specific model. There is much detail we have not addressed, such as the exact process by which the relevant statutory body would determine that the harm threshold had been breached, especially in borderline cases. Nor have we considered exemptions or administrative review. Crafting a single unified scheme to address cyber racism is an ambitious and controversial project. There will always be a degree of definitional uncertainty and public contestation that no legal instrument can hope to overcome in full. We can, however, attempt to combine legal and non-legal channels to better remedy cyber-racism and promote respectful online communities. Close monitoring of the effectiveness of the cyber-bullying scheme has a valuable contribution to make to the development of a framework suitable for achieving this goal. The AHRC should require intermediaries to provide public monitoring of the extent of and their responses to complaints on cyber racism. This data should be publicly available, so that decisions on complaints can produce a public data base of how the rules are applied in practice. Such an approach follows growing international best practice.

5. Moving forward

---

[212] *Broadcasting Services Act 1992* (Cth) pt 9.
[213] For example, if ACMA is satisfied that a person is providing subscription radio services other than in accordance with the relevant class license, ACMA may apply to the Federal Court for an order that the person cease providing those services: *Broadcasting Services Act 1992* (Cth) s 144. ACMA may apply to the Federal Court to enforce various civil penalty orders: see *Broadcasting Services Act 1992* (Cth) ss 205F-205G.
[214] In April 2016, the Federal government launched 'eSafetyWomen', an initiative of the Office of the Children's eSafety Commissioner (https://www.esafety.gov.au/women): Mitch Fifield, Minister for Communications, and Michaela Cash, Minister for Employment and Women, 'New eSafety Women Website Launched' (Media Release, 28 April 2016) <http://www.mitchfifield.com/Media/MediaReleases/tabid/70/articleType/ArticleView/articleId/1150/JOINT-MEDIA-RELEASE--New-eSafety-Women-website-launched.aspx>
[215] See Explanatory Memorandum, Enhancing Online Safety for Children Amendment Bill 2017 (Cth).

Just as government cannot afford to vacate the space of cyber-crime or cyber-bullying, neither can they afford to ignore cyber-racism. Given the ubiquity of online communication, the difficulties and sensitivities around regulating the Internet are no longer sufficient rationales for governments to take a 'hands off' approach. Any lack of will to regulate is resoundingly countered by evidence showing support for careful and tailored intervention. The reduction of harm associated with cyber racism requires close and continuing collaboration between government agencies, the Internet industry and civil society organisations. The Racism Stops with Me alliance could provide an umbrella for civil society to collaborate in this process of harm reduction and the building of resilience. The Commission is well placed to promote a public debate on these issues, and develop a national alliance on combatting cyber racism, in conjunction with industry partners, VicHealth, FECCA and the Online Hate Prevention Institute. A key regulatory issue remains the public availability of data that records the outcomes of all sources of intervention.

Cyber-racism poses a double challenge for regulators: ambiguity and controversy over legal definitions of racial speech; and amplification of regulatory difficulties on the Internet, including anonymity, dissemination and enforcement. There is a gap in current regulatory mechanisms to provide a prompt, efficient and enforceable system for denouncing and responding to the specific harm of racism in the digital environment.

In considering how to address this lacuna in protection, we propose a multi-pronged approach that places greater responsibility on industry codes of conduct, reinforced by state intervention, and public data availability. All legislative reform creates definitional debate. This uncertainty should not be seen as a fundamental obstacle to reform.
