REGULATING CYBER-RACISM

GAIL MASON* AND NATALIE CZAPSKI†

Cyber-racism and other forms of cyber-bullying have become an increasing part of the internet mainstream, with 35% of Australian internet users witnessing such behaviour online. Cyber-racism poses a double challenge for effective regulation: a lack of consensus on how to define unacceptable expressions of racism; and the novel and unprecedented ways in which racism can flourish on the internet. The regulation of racism on the internet sits at the crossroads of different legal domains, but there has never been a comprehensive evaluation of these channels. This article examines the current legal and regulatory terrain around cyber-racism in Australia. This analysis exposes a gap in the capacity of current regulatory mechanisms to provide a prompt, efficient and enforceable system for responding to harmful online content of a racial nature. Drawing on recent legislative developments in tackling harmful content online, we consider the potential benefits and limitations of key elements of a civil penalties scheme to fill the gap in the present regulatory environment. We argue for a multifaceted approach, which encompasses enforcement mechanisms to target both perpetrators and intermediaries once in-platform avenues are exhausted. Through our proposal, we can strengthen the arsenal of tools we have to deal with cyber-racism.

CONTENTS

* LLB (Qld), DipCrim (Melb), MA (Rutgers), PhD (La Trobe); Professor of Criminology, The University of Sydney. This research is supported by the Australian Research Council (LP120200115) and is part of a larger project on cyber-racism and community resilience. We would like to thank Andre Oboler, CEO, Online Hate Prevention Institute, for his advice on this research and Andrew Dyer for research assistance. We would also like to thank Alistair MacGibbon, the former Children's eSafety Commissioner, and Sharon Trotter, Manager Online Content, Office of the eSafety Commissioner, for their valuable insights.

† BA, LLB (Hons) (Syd); Research Assistant, Cyber-Racism and Community Resilience Project.

I Introduction

When retailer David Jones appointed Aboriginal Australian Adam Goodes as a brand ambassador in October 2015, they were perhaps unprepared for the deluge of online animosity that flooded their Facebook page.[1] Goodes, a retired Australian Football League ('AFL') player, had already been the target of racist remarks on the field. It could be suggested that much of the online abuse was deemed too repugnant to be published by the commentators who documented the incident.[2] While there is nothing new about public expressions of racial hostility, the comments directed at Goodes provide one illustration of the proliferation of such hostility through online means.

Much has been written in Australia about the regulation of offline vilification based on race, colour, national or ethnic origin and religion.[3] We know far less about how the law tackles online vilification. This article examines the current state of regulation in Australia as it relates to 'cyber-racism', specifically legal, quasi-legal and self-regulatory regimes.

We begin, in Part II, by arguing that cyber-racism presents a double challenge for regulation. First is the difficulty of defining racism itself, including the absence of consensus about where the boundary lies between tolerable and intolerable racial speech as well as the lack of a clear demarcation between speech based on race, ethnicity, nationality, religion and the like. Second is the challenge of policing the digital environment. New technologies enable and expand the avenues for racial vilification, raising distinct problems of dissemination, anonymity and enforcement. In Australia, the regulation of racism on the internet sits at the crossroads of several different legal domains.

1 Leesha McKenny, 'David Jones Flooded with Abuse after Adam Goodes Announced as Ambassador', *The Sydney Morning Herald* (Sydney, 19 October 2015) <www.smh.com.au/business/retail/david-jones-flooded-with-abuse-after-adam-goodes-announced-as-ambassador-20151018-gkcacz.html>, archived at <https://perma.cc/4X3G-QR78>.

2 See, eg, Lucy Mae Beers, '"I Will Not Spend Another Cent There": David Jones Facebook Page Inundated with Racist Abuse after Retail Giant Appoints Adam Goodes Brand Ambassador', *Daily Mail Australia* (Sydney, 19 October 2015) <www.dailymail.co.uk/news/article-3278693/David-Jones-Facebook-page-inundated-racist-abuse-boycott-threats-appoints-Adam-Goodes-ambassador.html>.

3 See, eg, Luke McNamara, *Regulating Racism: Racial Vilification Laws in Australia* (Institute of Criminology, 2002); Katharine Gelber and Adrienne Stone (eds), *Hate Speech and Freedom of Speech in Australia* (Federation Press, 2007); Margaret Thornton and Trish Luker, 'The Spectral Ground: Religious Belief Discrimination' (2009) 9 *Macquarie Law Journal* 71; Katharine Gelber and Luke McNamara, 'Private Litigation to Address a Public Wrong: A Study of Australia's Regulatory Response to "Hate Speech"' (2014) 33 *Civil Justice Quarterly* 307.

There has never been a comprehensive evaluation of these channels. For this reason, our focus in this article is more squarely on this second challenge.

Part III examines the current legal and regulatory terrain in relation to cyber-racism in Australia, with reference to vilification laws and attendant conciliation schemes, the criminal law, the *Broadcasting Services Act 1992* (Cth) ('*BSA*') and the new Commonwealth cyber-bullying legislation, as well as intermediary[4] terms of service and codes of conduct. Despite the available spectrum of civil, criminal and voluntary avenues, we conclude in Part IV that there is a significant gap in the regulatory environment in Australia. There is no comprehensive system for expressly denouncing and remedying the harm of cyber-racism by offering an efficient and accountable process for removing harmful material, backed by a mechanism of enforcement.

Keeping debates around freedom of speech and freedom from racism firmly in mind,[5] in Part V we draw out key elements of the recent civil penalties scheme for cyber-bullying in Australia[6] to explore the utility of this approach for addressing cyber-racism. We conclude that the range of views about how to define and respond to racist speech is best recognised through a multi-pronged approach that places greater regulatory responsibility on internet intermediaries, while also offering aggrieved parties effective and enforceable avenues for confronting speech they find intolerable.

II The Double Challenge of Cyber-Racism

A *Defining Racism and Racist Speech*

There is evidence that racism is a significant problem in Australia, both offline and online. In 2016, 20% of Australians experienced discrimination based on 'skin colour, ethnic origin or religion', nearly a third of those 'about once

4 Danielle Citron and Helen Norton use the term 'intermediary' to refer to 'private entities that host or index online content', 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 *Boston University Law Review* 1435, 1438–9.

5 See, eg, Peter Wertheim, 'Freedom and Social Cohesion: A Law that Protects Both' (Conference Paper, 40 Years of the *Racial Discrimination Act 1975* (Cth) Conference, 19–20 February 2015) <www.humanrights.gov.au/our-work/race-discrimination/publications/perspectives-racial-discrimination-act-papers-40-years>, archived at <https://perma.cc/4PQU-7HZS>.

6 *Enhancing Online Safety Act 2015* (Cth).

a month' or 'most weeks in the year'.[7] On the basis of surveys conducted over the past three decades, the Scanlon Foundation concluded in 2014 that the 'core level of intolerance in Australia … [amounted to nearly] 10% of the population'.[8]

A preliminary complexity in crafting a legal response to racism is the question of how to define it. The absence of a universally accepted definition should come as no surprise given that racial categories themselves are social constructs without firm empirical foundation.[9] Still, the concept of race is employed in contemporary policy and legal instruments to designate perceived differences between groups of people based on physical and social characteristics, such as skin colour, ethnicity and national origin.[10] It follows that racism has come to be used as an umbrella term to refer to a combination of values, attitudes and behaviours that exclude people from society on the basis of their race, ethnicity, cultural practices, national origins, indigeneity and, in some instances, religious beliefs.[11]

7 Andrew Markus, *Mapping Social Cohesion* (Scanlon Foundation Surveys, 2016) 25–6. This was up from 15% in 2015, 18% in 2014, 19% in 2013, and well above the 9% reported by the Scanlon Foundation in 2007: Andrew Markus, *Mapping Social Cohesion* (Scanlon Foundation Surveys, 2014) 3, 19 ('2014 Survey'); Andrew Markus, *Mapping Social Cohesion* (Scanlon Foundation Surveys, 2015) 23. See also Kevin Dunn et al, 'Cities of Race Hatred? The Spheres of Racism and Anti-Racism in Contemporary Australian Cities' (2009) 1(1) *Cosmopolitan Civil Societies Journal* 1, 1.

8 Markus, 2014 Survey (n 7) 58.

9 Robert Miles and Malcolm Brown, *Racism* (Routledge, 2nd ed, 2003); Yin C Paradies, 'Defining, Conceptualizing and Characterizing Racism in Health Research' (2006) 16 *Critical Public Health* 143, 144.

10 OSCE Office for Democratic Institutions and Human Rights and the International Association of Prosecutors, *Prosecuting Hate Crimes* (Practical Guide, 2014) 29, 30. Many contemporary definitions of race include ethnicity which, in turn, denotes characteristics that include 'religion, culture, geographical origin, history and language': at 30. The definition of racial discrimination also includes reference to 'race, colour, descent or national or ethnic origin': *International Convention on the Elimination of All Forms of Racial Discrimination*, opened for signature 21 December 1965, 660 UNTS 212 (entered into force 4 January 1969) art 1; *Racial Discrimination Act 1975* (Cth) s 9 ('*RDA*').

11 See Y Paradies et al, *Building on Our Strengths: A Framework to Reduce Racial Discrimination and Promote Diversity in Victoria* (Report, 2009) 7 <http://ro.uow.edu.au/scipapers/4674/>, archived at <https://perma.cc/92CT-BESZ>; Andrew Jakubowicz, 'Hunting for the Snark and Finding the Boojum: Building Community Resilience against Race Hate Cyber Swarms' (Conference Paper, 40 Years of the *Racial Discrimination Act 1975* (Cth) Conference, 19–20 February 2015) 105, 106–8 <www.humanrights.gov.au/our-work/race-discrimination/publications/perspectives-racial-discrimination-act-papers-40-years>, archived at <https://

Racist speech, whether through language, images or symbols, is a tangible manifestation of racism. Its regulation is said to be justified on the grounds that it causes significant individual harm to the recipient's sense of dignity, wellbeing and safety,[12] as well as group harm to the target community who may interpret such expressions as a sign of intolerance and victimisation.[13] Racism is also said to embody a moral failure to treat others equally, decently and fairly.[14] Its regulation stands as a public denunciation of attitudes that undermine the values of multiculturalism and equality implicit in liberal democracies.[15] This violation of shared values makes racist speech a 'public wrong'.[16]

This understanding of racist speech belies two further complexities. First is the application of the term 'racism' to expressions of bias towards intersecting characteristics, such as ethnicity, language, nationality, tribal linkages, immigration status or religion. Without downplaying the distinct aetiologies of different forms of prejudice, the line between these bases of discrimination can be porous.[17] For example, Arab Australians report that discrimination directed towards them may be based on the assumption that they are Mus-

perma.cc/Z8A8-9LD5>. Although legislators tend to prefer to use the term 'racial' over 'racism', Goodall argues that 'racial' implies the same moral fault inherent in 'racism': Kay Goodall, 'Conceptualising "Racism" in Criminal Law' (2013) 33 *Legal Studies* 215, 218, 232, 234.

12 Paradies et al (n 11) 36. See also Gabrielle Berman and Yin Paradies, 'Racism, Disadvantage and Multiculturalism: Towards Effective Anti-Racist Praxis' (2010) 33 *Ethnic and Racial Studies* 214, 215–18.

13 See Sentencing Advisory Council, *Sentencing for Offences Motivated by Hatred or Prejudice* (Report of Advice, July 2009) 1 [A.4], quoting Manitoba Department of Justice, *Policy Directive: Hate Motivated Crime* (Guideline No 2:HAT:1, June 2008) 5 <www.gov.mb.ca/justice/prosecutions/pubs/hate_crimes.pdf>, archived at <https://perma.cc/YF4Z-FAGW>.

14 Tim Soutphommasane, 'Racism is a Moral Issue' (Speech, Society of Australasian Social Psychologists Conference, 11 April 2014) <www.humanrights.gov.au/news/speeches/racism-moral-issue>, archived at <https://perma.cc/YV4C-T8YT>.

15 See Frederick M Lawrence, 'Enforcing Bias-Crime Laws without Bias: Evaluating the Disproportionate-Enforcement Critique' (2003) 66(3) *Law and Contemporary Problems* 49, 51. See also Paradies et al (n 11) 45–6.

16 Gelber and McNamara, 'Private Litigation to Address a Public Wrong' (n 3) 309.

17 Thornton and Luker (n 3) 91. The authors point out that perceptions about the acceptability of speech are also shaped by differences in the degree to which a particular attribute is said to be innate or chosen.

lim,[18] creating a 'blurred line between race and religion for victims of racism'.[19] With some exceptions, the weight of Australian and international authority gives a fairly broad interpretation to the concept of race as one that should be 'used in a popular sense' rather than restricted to biological tests.[20] We revisit the legal technicalities of this issue in more detail later in this article but suffice to say here that the concept of racist speech that we employ in this article is comparably broad. It avoids the historical inaccuracy of rigid biological accounts of race by encompassing categories of speech that bleed into more 'popular' interpretations of racism, including those based on ethno-religion, skin colour, language, culture and national origin.

Second is the difficulty of drawing a distinction between racial speech that warrants prohibition and that which does not or, to put this in another way, the question of how to strike a balance between freedom of expression and protection from racist speech.[21] Liberal democracies have taken a variety of nuanced approaches, which are grounded in different historical, political and institutional contexts.[22] At one end of the spectrum is the United States, which has bucked the postwar trend towards limiting freedom of expression to

18 Human Rights and Equal Opportunity Commission, *Isma*ξ *— Listen: National Consultations on Eliminating Prejudice against Arab and Muslim Australians* (Report, 2003) 45 <www.humanrights.gov.au/our-work/race-discrimination/projects/isma-listen-national-consultations-eliminating-prejudice>, archived at <https://perma.cc/Y9KZ-65T2>. See generally Thornton and Luker (n 3) 91.

19 Kate Eastman, 'Mere Definition? Blurred Lines? The Intersection of Race, Religion and the *Racial Discrimination Act 1975* (Cth)' (Conference Paper, 40 Years of the *Racial Discrimination Act 1975* (Cth) Conference, 19–20 February 2015) 147 <www.humanrights.gov.au/our-work/race-discrimination/publications/perspectives-racial-discrimination-act-papers-40-years>, archived at <https://perma.cc/2D46-4RQL>. While Jews have been recognised as an ethnic group and covered by the *RDA*, Muslims have not. See also Thornton and Luker (n 3) 79.

20 *King-Ansell v Police* [1979] 2 NZLR 531, 542 (Richardson J). See also *Ealing London Borough Council v Race Relations Board* [1972] AC 342, 362 (Lord Simon); *Williams v Tandanya Cultural Centre* (2001) 163 FLR 203, 209 [21].

21 Australian Human Rights Commission ('AHRC'), 'Human Rights in Cyberspace' (Background Paper, September 2013) 14–15. See generally Erik Bleich, *The Freedom to Be Racist? How the United States and Europe Struggle to Preserve Freedom and Combat Racism* (Oxford University Press, 2011).

22 Bleich (n 21). Erik Bleich draws upon examples from a number of jurisdictions, including the United States, Denmark, Germany, France and the United Kingdom to demonstrate the nuanced approaches and challenges associated with regulating racist speech.

protect minority groups by penalising racist speech.[23] Although not without limits, the United States has extensively entrenched protections around the freedom to express offensive and provocative speech, underwritten by interpretation given to the First Amendment.[24] In contrast, European jurisdictions have slowly but consistently expanded human rights protection through restrictions on the expression of racist views over the past few decades.[25] This is exemplified by France, which takes a strong legislative approach, and has had a number of high profile prosecutions for racist speech in recent years.[26]

Australia has racial vilification laws in place in almost every jurisdiction[27] but there continues to be a divergence of views about where to set the legal threshold between acceptable and unacceptable speech.[28] Recently, this debate was reignited with the establishment of a Parliamentary Inquiry into Freedom of Speech in Australia. This examined, amongst other things, the operation of s 18C of the *Racial Discrimination Act 1975* (Cth) ('*RDA*').[29] It is unlikely that

23 Ibid 6–7. Notably, however, Bleich points out that although racists in the United States are 'free to think and say almost anything', such freedom ends as soon as racists 'act upon their beliefs', noting that the US has some of the most prolific legislation around discrimination and hate crime: at 7.

24 Ibid 62–3. See Citron and Norton (n 4) 1438–9; Andre Oboler and Karen Connelly, 'Hate Speech: A Quality of Service Challenge' (Conference Paper, IEEE Conference on e-Learning, e-Management and e-Services (IC3e), 10–12 December 2014) 118.

25 Bleich (n 21) 17.

26 Ibid 17, 29. Examples of this include French former film star Brigitte Bardot, who has been convicted five times on charges of inciting racial hatred: at 17. See also 'Ex-Film Star Bardot Gets Fifth Racism Conviction', *Reuters India* (Mumbai, 3 June 2008) <http://in.reuters.com/article/idINIndia-33883520080603>, archived at <https://perma.cc/EHW8-H4E7>. Marine Le Pen, leader of the French far right National Front Party, was acquitted of racial hatred charges in December 2015: see Noemie Bisserbe, 'Marine Le Pen Acquitted of Inciting Racial Hatred', *The Wall Street Journal* (New York City, 15 December 2015) <www.wsj.com/articles/marine-le-pen-acquitted-of-inciting-racial-hatred-1450193282>.

27 The Northern Territory is the only Australian jurisdiction without racial vilification laws: see Thornton and Luker (n 3) 84.

28 For example, although the bulk of submissions to the 2014 Commonwealth Attorney-General's consultation on proposed amendments to s 18C of the *RDA* were in favour of the existing laws, a minority of views were in favour of change: see Jakubowicz (n 11) 106–8. Section 18C of the *RDA* makes it a civil wrong for a person 'to do an act, otherwise than in private, … [which is] reasonably likely, in all the circumstances, to offend, insult, humiliate or intimidate another person or a group' on the basis of their 'race, colour or national or ethnic origin'.

29 George Brandis, 'Parliamentary Inquiry into Freedom of Speech' (Media Release, 8 November 2016) <www.attorneygeneral.gov.au/Mediareleases/Pages/2016/FourthQuarter/

any legal standard will completely resolve this question which, as we discuss below, calls for a combination of legal and non-legal responses. It is notable, however, that a 2014 Nielsen survey shows that 88% of Australians believe it should be unlawful to offend, insult or humiliate others on the basis of race.[30] This support amongst the Australian public for formal regulation is testament to law's importance in adjudicating a path between the differences of opinion that surround state-based sanctions of racial speech.

B *Defining Cyber-Racism*

Law's role in regulating race-based speech now extends to the virtual world. In its early days, some scholars imagined the internet as the ultimate space of democratisation,[31] where individuals could escape from racial markers and racism.[32] And yet, far from being a colour-blind space, expressions of race (both positive and negative) have only proliferated online.[33] 'Web 2.0' technologies, such as Facebook and Twitter, video-sharing platform YouTube, and various blogging and community platforms produce vast amounts of user-generated content, providing new avenues for the dissemination of racial

Parliamentary-inquiry-into-freedom-of-speech.aspx>, archived at <https://perma.cc/U5BG-9BL5>. This inquiry follows the Federal Court dismissing a case against three Queensland University of Technology students: see Jane Norman, '18C Inquiry on the Cards as Malcolm Turnbull Softens Position on Amending Racial Discrimination Act', *ABC News* (Sydney, 8 November 2016) <www.abc.net.au/news/2016-11-07/18c-inquiry-on-the-cards-malcolm-turnbull-confirms/8001292>, archived at <https://perma.cc/74FH-6F34>. The Committee reported to Parliament on 28 February 2017: Parliamentary Joint Committee on Human Rights, *Inquiry into the Operation of Part IIA of the Racial Discrimination Act 1975 (Cth) and Related Procedures under the Australian Human Rights Commission Act 1986 (Cth)* (Inquiry Report, 28 February 2017).

30 'Overwhelming Majority Reject Change to Racial Vilification Law', *Australian Human Rights Commission* (Web Page, 14 April 2014) <www.humanrights.gov.au/news/stories/overwhelming-majority-reject-change-racial-vilification-law>, archived at <https://perma.cc/8GD2-DGBS>.

31 Citron and Norton (n 4) 1443–6.

32 Henry Jenkins, 'Cyberspace and Race: The Color-Blind Web', *MIT Technology Review* (Cambridge, Massachussetts, 1 April 2002) <www.technologyreview.com/article/401404/cyberspace-and-race/>, archived at <https://perma.cc/Q7HB-LFAS>; Lisa Nakamura and Peter A Chow-White, 'Introduction: Race and Digital Technology' in Lisa Nakamura and Peter A Chow-White (eds), *Race after the Internet* (Routledge, 2012) 1, 17; Jessie Daniels, 'Race and Racism in Internet Studies: A Review and Critique' (2013) 15 *New Media and Society* 695, 695.

33 Nakamura and Chow-White (n 32) 5; Daniels (n 32).

commentary.[34] Facebook, for example, has over 1 billion daily active users,[35] and YouTube has over 1 billion users.[36]

Cyber-racism refers to racism that manifests in this online world. It includes words, images and symbols posted on social media services,[37] online games, forums, messaging services and dedicated 'hate sites'.[38] Cyber-racism includes a wide spectrum of conduct in terms of seriousness and specificity, ranging from, for example, racist material disguised as 'humour'[39] to direct threats and incitements to violence targeting specific individuals or groups on the basis of race.[40]

Adopting a broad understanding of the concept of racism, recent research shows that nearly 35% of Australian internet users have witnessed cyber-

34 Yaman Akdeniz, *Racism on the Internet* (Council of Europe Publishing, 2009) 14; Abraham H Foxman and Christopher Wolf, *Viral Hate: Containing its Spread on the Internet* (Palgrave Macmillan, 2013) 11.

35 'Company Info: Stats', *Facebook Newsroom* (Web Page, 2017) <http://newsroom.fb.com/company-info/>, archived at <https://perma.cc/97LN-5TSM>.

36 'YouTube by the Numbers', *YouTube for Press* (Website, 2017) <www.youtube.com/yt/press/statistics.html>, archived at <https://perma.cc/3YFZ-GBLX>. YouTube users also extensively comment on uploaded videos, generating a further mass of online content.

37 This article will use the term 'social media services/platforms' widely to include 'technologies that enable the production and sharing of digital content in mediated social settings', where users can interact with others and share material: Danielle Keats Citron, 'Fulfilling Government 2.0's Promise with Robust Privacy Protections' (2010) 78 *George Washington Law Review* 822, 824 n 12; Citron and Norton (n 4) 1439 n 22. Note also the broad definition of 'social media service' adopted by the *Enhancing Online Safety Act 2015* (Cth) s 9(1).

38 Brendesha M Tynes et al, 'Online Racial Discrimination and the Protective Function of Ethnic Identity and Self-Esteem for African American Adolescents' (2012) 48 *Developmental Psychology* 343, 344. Tynes et al and some others include text-messaging services within their definition of online platforms through which racism may be perpetrated. This is acknowledged, but is not a focus of the present article. See also Imran Awan, 'Islamophobia and Twitter: A Typology of Online Hate against Muslims on Social Media' (2014) 6 *Policy and Internet* 133, 134, 139.

39 Simon Weaver, 'Jokes, Rhetoric and Embodied Racism: A Rhetorical Discourse Analysis of the Logics of Racist Jokes on the Internet' (2011) 11 *Ethnicities* 413, 431. Weaver argues that 'humour can act as a form of racist rhetoric for serious racism and thus should not always be seen as "just a joke" or fundamentally harmless'.

40 The barrage of hate directed at Australian Muslim and activist lawyer Mariam Veiszadeh is a prominent recent example of this: Kim Stephens, 'Mariam Veiszadeh Now the Target of US Anti-Islam Site', *The Queensland Times* (Ipswich, Queensland, 25 February 2015) <www.qt.com.au/news/mariam-veiszadeh-now-target-us-anti-islam-site/2555598/>, archived at <https://perma.cc/RE95-TD4U>.

racism.[41] The main targets are Indigenous Australians, Middle Eastern people, Africans, Muslims and Jews.[42] In recent years, a substantial proportion of racial hatred complaints to the Australian Human Rights Commission ('AHRC') have concerned internet material, peaking in 2012–13 with 41% of complaints about conduct online.[43] The US-based Simon Wiesenthal Center reported over 14,000 'hate speech' websites in 2012, and the number of extremist websites appears to be on the rise.[44] Although there is relatively limited evidence on how racism online affects its targets, it is apparent that, like racism in the offline world, online racism can negatively affect self-esteem and produce feelings of anger, frustration and hopelessness.[45] It is also related to higher levels of depression and anxiety.[46] Whilst some persons who engage in cyber-racism are members of extremist groups, many others are 'ordinary' Australian citizens expressing prejudicial views.[47]

41 Karen Connelly, 'Understanding Cyber-Racism and Building Community Resilience', *Australian Mosaic* (Australian Capital Territory, December 2015) 43.

42 This is demonstrated by preliminary results (unpublished) from the 'Encounters' stream of the ARC Cyber-Racism and Community Resilience project, of which the current research is a part. The Encounters stream surveyed 2,141 participants through two online MyOpinion panels, targeting the general population as well as groups significantly at risk of racism: see Kevin Dunn, Yin Paradies and Rosalie Atie, *Preliminary Result: Cyber Racism and Community Resilience the Survey* (Research Report, 28–29 May 2014) 3.

43 The percentage of racial hatred complaints categorised under the sub-area 'Internet — email/webpage/chatroom' has fluctuated in recent years: 34% in 2009–10; 23% in 2010–11; 17% in 2011–12; 41% in 2012–13; 10% in 2013–14; 8% in 2014–15): AHRC, *Annual Report 2009–2010* (Report, 30 September 2010) 80; AHRC, *Annual Report 2010–2011* (Report, 30 September 2011) 109; AHRC, *Annual Report 2011–2012* (Report, 28 September 2012) 134; AHRC, *Annual Report 2012–2013* (Report, 8 October 2013) 133; AHRC, *Annual Report 2013–2014* (Report, 4 March 2015) 139; AHRC, *Annual Report 2014–2015* (Report, 12 October 2015) 143.

44 Alisdair A Gillespie, *Cybercrime: Key Issues and Debates* (Routledge, 2015) 183.

45 Dunn, Paradies and Atie (n 42) 13.

46 Tynes et al, 'Online Racial Discrimination and the Protective Function of Ethnic Identity and Self-Esteem for African American Adolescents' (n 38) 345, citing Brendesha M Tynes et al, 'Online Racial Discrimination and Psychological Adjustment among Adolescents' (2008) 43 *Journal of Adolescent Health* 565.

47 Survey research conducted in 2006 in Sydney, Melbourne and Perth found that between 1 in 10 and 1 in 3 respondents experienced some form of 'everyday racism', depending upon their background and situation, such as 'racist talk' (verbal abuse, name calling or racist slurs), exclusion, unfair treatment, or attack: Dunn et al (n 7) 1, 5. The National Inquiry into Racist Violence in Australia, conducted in 1991, also pointed to the 'everyday' nature of violence, in finding that a high proportion of racist violence was carried out by neighbours: see especially

Cyber-racism presents its own specific regulatory challenges. Material posted online is ubiquitous and relatively permanent. Information can be disseminated instantaneously, continuously and globally, reaching far greater audiences than practicable in the offline world.[48] Moreover, when material is published online, it remains '"cached" or stored', and can potentially be accessed via search engines, and easily duplicated.[49] The removal of harmful content from a platform does not guarantee its erasure from cyberspace. This has led some to describe cyber-hate as a 'permanent disfigurement' on members of the targeted group.[50]

The online world connects people with real or imagined communities of others who share their viewpoints. This adds credibility to those who already harbour discriminatory views.[51] It emboldens users to express racist views and makes them less willing to compromise.[52] This can fuel a 'mob-like' approach to harassment of victims.[53] Of course, the vast amounts of information gathered online can help enforcement by making it easier to identify some perpetrators.[54] However, the use of 'anonymous' profiles and pseudonyms,

Irene Moss and Ron Castan, *Racist Violence: Report of the National Inquiry into Racist Violence in Australia* (Report, 27 March 1991) 374–5.

48 New Zealand Law Commission, 'Harmful Digital Communications: The Adequacy of the Current Sanctions and Remedies' (Ministerial Briefing Paper, August 2012) 10; AHRC, 'Human Rights in Cyberspace' (n 21) 16. Australian courts have commented on a number of occasions on the ubiquity and accessibility of internet material and the difficulties this poses for the administration of justice: see, eg, *General Television Corporation Pty Ltd v DPP (Vic)* (2008) 19 VR 68, 88 [70].

49 AHRC, 'Human Rights in Cyberspace' (n 21) 16.

50 Citron and Norton (n 4) 1452, citing Jeremy Waldron, 'Dignity and Defamation: The Visibility of Hate' (2010) 123 *Harvard Law Review* 1597, 1601, 1610.

51 Larry Keller, 'Experts Discuss the Role of Race Propaganda after White Massachusetts Man Kills Two African Immigrants' [2009] (Summer Issue) *Intelligence Report* (online) <www.splcenter.org/fighting-hate/intelligence-report/2009/experts-discuss-role-race-propaganda-after-white-massachusetts-man-kills-two-african>, archived at <https://perma.cc/VN8Y-ZDRS>.

52 See Anne Pedersen, Brian Griffiths and Susan E Watt, 'Attitudes toward Out-Groups and the Perception of Consensus: All Feet Do *Not* Wear One Shoe' (2008) 18 *Journal of Community and Applied Social Psychology* 543, 554.

53 Danielle Keats Citron, *Hate Crimes in Cyberspace* (Harvard University Press, 2014) 5.

54 For example, a chiropractor based in Woy Woy who was recently convicted of posting abusive and racist comments online about the former Senator Nova Peris was identified through his Facebook account: see 'Nova Peris Accepts Apology from Disgraced Chiropractor Chris Nelson', *ABC News* (Sydney, 13 July 2016) <www.abc.net.au/news/2016-07-

quite apart from having a disinhibiting effect on perpetrators,[55] makes it difficult for victims of hateful conduct to seek informal redress.[56] At times, legal compulsion of internet providers/content hosts has been necessary to reveal the identities of perpetrators.[57]

Further, the internet often bypasses the traditional media gatekeepers that act as a check on the dissemination of unpalatable viewpoints,[58] with content easily spread without regard for state boundaries. The cross-jurisdictional nature of online 'publication' makes it possible to have an Australian victim, targeted by a cyber-racist in another country, on a social media platform hosted by a third company incorporated in a fourth jurisdiction. Dealing with any one instance of cyber-racism may require coordination between law enforcement and government agencies from multiple countries as well as intermediaries such as online host platforms and connectivity providers, bringing to light legal inconsistencies between jurisdictions.[59]

13/nova-peris-accepts-chiropractor's-apology/7627240>, archived at <https://perma.cc/X4CJ-W6YR>.

55 AHRC, 'Human Rights in Cyberspace' (n 21) 17. According to Bocij and McFarlane, a combination of social and technological factors encourage citizens to partake in anti-social and criminal behaviours online, in which they might never engage in the offline world: Paul Bocij and Leroy McFarlane, 'Cyberstalking: The Technology of Hate' (2003) 76 *Police Journal* 204, 207–8.

56 AHRC, 'Human Rights in Cyberspace' (n 21) 18–19. However, Citron argues that removing anonymity is not the answer to combating online harassment as determined harassers 'can easily work around' the restrictions, whilst others for whom anonymity is needed as a protection from abuse will be silenced: Citron, *Hate Crimes in Cyberspace* (n 53) 239.

57 One example of this is a 2012 UK High Court decision which ordered Facebook to reveal the names, email addresses and IP addresses of users who systematically abused a British woman: Vanessa Allen, 'Victory over Cyber Bullies: Legal First as High Court Orders Facebook to Reveal Trolls Who Tormented Mother for Defending X Factor Star', *Daily Mail Australia* (Sydney, 8 June 2012) <www.dailymail.co.uk/news/article-2156365/Nicola-Brookes-victim-internet-trolls-wins-High-Court-backing-reveal-identities-targeted-her.html>. See also AHRC, 'Human Rights in Cyberspace' (n 21) 17.

58 Majid Yar, *Cybercrime and Society* (Sage Publications, 2006) 102.

59 Facebook's data policy, for example, states: 'We may access, preserve and share your information in response to a legal request (like a search warrant, court order or subpoena) if we have a good faith belief that the law requires us to do so. This may include responding to legal requests from jurisdictions outside of the United States where we have a good faith belief that the response is required by law in that jurisdiction, affects users in that jurisdiction, and is consistent with internationally recognized standards': 'Data Policy', *Facebook* (Web Page, 29 September 2016) <www.facebook.com/full_data_use_policy>, archived at <https://perma.cc/F5H6-UP9V>.

C *Conclusion: The Double Challenge of Cyber-Racism*

In some ways, the online expression of racism is simply an extension of racist conduct that occurs in the physical world, bringing with it the same challenges and controversies that lie at the core of all anti-racism policy and law. However, the internet provides unprecedented and novel opportunities for racism to flourish.[60] The sheer volume of material and the speed of its dissemination to a 'wider audience than was ever possible before'[61] means that isolated events and commentary can have global effects.[62] Traditional media regulations may be easily bypassed under the cover of anonymity and the unmediated nature of the online environment.

In effect, regulating cyber-racism presents a double challenge: (i) the complex problem of defining illegal speech on the basis of race; and (ii) the difficulties of policing the internet. In Australia, this first challenge is well recognised[63] and it is not the aim of this article to resolve it. Conversely, we do not have a comprehensive picture of the regulatory channels available in Australia to respond to racist speech on the internet, or their limitations. The following section will provide such an overview, examining both legal and non-legal avenues.

III The Current Legal and Regulatory Terrain

A *Federal and State/Territory Racial Vilification Laws*

1 *Civil Racial Vilification Laws*

At the federal level, s 18C of the *RDA* 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 or national

60 Daniels (n 32) 695–6.

61 Kevin M Dunn and Rosalie Atie, 'Regulating Online Racism in the Online Age' (Conference Paper, 40 Years of the *Racial Discrimination Act 1975* (Cth) Conference, 19–20 February 2015) 118 <www.humanrights.gov.au/our-work/race-discrimination/publications/perspectives-racial-discrimination-act-papers-40-years>, archived at <https://perma.cc/CQG7-3NZM>.

62 Brian McNair, 'When Terror Goes Viral It's Up to Us to Prevent Chaos', *The Conversation* (Melbourne, 27 July 2016) <https://theconversation.com/when-terror-goes-viral-its-up-to-us-to-prevent-chaos-62687>, archived at <https://perma.cc/J4LP-9ZVX>.

63 See n 3.

or ethnic origin'.[64] The act must be done 'otherwise than in private',[65] which has been interpreted to include conduct occurring online, such as material published on a website that 'is not password protected'.[66]

There is also racial vilification legislation in every state and territory, with the exception of the Northern Territory, intended to operate concurrently with Commonwealth laws.[67] While most jurisdictions have both civil and criminal provisions, Tasmania has only a civil prohibition,[68] and Western Australia deals with racial vilification only through the criminal law.[69] The state and territory civil laws are largely based upon the NSW vilification legislation, which, as Rees, Rice and Allen describe, 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'.[70] 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,[71] a 'public act' is broadly defined in the NSW legislation to include 'any form of communication to the public'.[72] It has been

64 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* (n 10).

65 *RDA* (n 10) s 18C(1).

66 *Jones v Toben* (2002) 71 ALD 629, 646 [74], aff'd (2003) 129 FCR 515.

67 In the Northern Territory, ss 19(1)(r) and 28 of the *Anti-Discrimination Act 1992* (NT) prohibit discrimination in the course of education, employment, accommodation, goods, clubs, and insurance and superannuation, or on the ground that a person associates with a person of aparticular race, sex, and/or age (amongst others characteristics). The Commonwealth vilification laws apply concurrently alongside state laws in the Northern Territory, as they do throughout Australia: *RDA* (n 10) s 18F; Neil Rees, Simon Rice and Dominique Allen, *Australian Anti-Discrimination Law* (Federation Press, 2nd ed, 2014) 617.

68 *Anti-Discrimination Act 1998* (Tas) ss 19–21.

69 *Criminal Code Act Compilation Act 1913* (WA) ss 77–80D.

70 Rees, Rice and Allen (n 67) 670. See also *Discrimination Act 1991* (ACT) s 67A; *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) s 19; *Racial and Religious Tolerance Act 2001* (Vic) s 7.

71 *Racial and Religious Tolerance Act 2001* (Vic) s 7.

72 *Anti-Discrimination Act 1977* (NSW) s 20B.

interpreted elsewhere to encompass the publication of material online.[73] 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.[74] The *RDA* does not operate extraterritorially.[75] 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 would fall within the bounds of the legislation.[76] A similar argument is likely to apply to state and territory vilification legislation.[77]

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.[78] It is sufficient to show that a particular subset of a racial group is reasonably likely to be affected by the conduct.[79] The conduct in question must cause 'profound and serious effects, not to be likened to mere slights'.[80] Conversely, the

73 In *Collier v Sunol* [2005] NSWADT 261, [33], material involving homosexual vilification published online, and publicly accessible, constituted a 'public act' under the *Anti-Discrimination Act 1977* (NSW).

74 *RDA* (n 10) s 18D; *Discrimination Act 1991* (ACT) s 67A(2); *Anti-Discrimination Act 1977* (NSW) s 20C(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. See also AHRC, 'Cyber Racism and Community Resilience Project: Civil and Criminal Racial Vilification Provisions' (Working Paper, July 2015) 8–10.

75 *Brannigan v Commonwealth* (2000) 110 FCR 566, 572–3 [26].

76 See AHRC, 'Cyber Racism and Community Resilience Project' (n 74) 26. See, eg, *Dow Jones & Company Inc v Gutnick* (2002) 210 CLR 575 ('*Gutnick*'), 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.

77 The Victorian civil and criminal vilification provisions expressly apply to conduct occurring outside Victoria: *Racial and Religious Tolerance Act 2001* (Vic) ss 7(2)(b), 24(3)(b).

78 *Creek v Cairns Post Pty Ltd* (2001) 112 FCR 352, 356 [13]; AHRC, 'Cyber Racism and Community Resilience Project' (n 74) 6. 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 multicultural and tolerant society.

79 AHRC, 'Cyber Racism and Community Resilience Project' (n 74) 6. See, eg, *McGlade v Lightfoot* (2002) 124 FCR 106, 117 [46]; *Eatock* (n 78) 363 [452].

80 *Creek* (n 78) 356 [16].

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.[81]

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.[82] This is difficult to prove and less satisfactory for the victim, being divorced from their own personal reactions[83] or any assessment of the respondent's motive or intention in performing the act.[84] 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.[85] 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.[86] There also continue to be gaps in the coverage afforded to religious vilification under both federal

81 In *Catch the Fire Ministries Inc v Islamic Council of Victoria Inc* (2006) 15 VR 207, 249 [132], 254–5 [158], Ashley JA and Neave JA held that the effect of the conduct under Victorian legislation should be assessed 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: at 212–13 [18]. In *Sunol v Collier [No 2]* (2012) 289 ALR 128, 136–7 [33]–[34], 137–8 [41], 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 *Veloskey v Karagiannakis* [2002] NSWADTAP 18, [28], where it was held that racial vilification legislation should consider the effect on the 'ordinary reasonable reader'.

82 Rees, Rice and Allen (n 67) 671; AHRC, 'Cyber Racism and Community Resilience Project' (n 74) 21.

83 AHRC, 'Cyber Racism and Community Resilience Project' (n 74) 22.

84 Rees, Rice and Allen (n 67) 671; AHRC, Submission to the Commonwealth Attorney-General's Department, *Amendments to Part IIA, Racial Discrimination Act* (28 April 2014) 21 [97]–[99].

85 AHRC, 'Cyber Racism and Community Resilience Project' (n 74) 21.

86 Ibid 22.

and state laws.[87] 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.[88]

In the majority of Australian jurisdictions, complaints of racial vilification are handled through a confidential process of conciliation wherever possible.[89] 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 may instigate a complaint in court;[90] equally, a civil case cannot be instigated unless a complaint to the AHRC has been terminated under certain provisions.[91]

2 *Criminal Racial Vilification Laws*

Most jurisdictions also have a criminal offence of 'serious racial vilification',[92] adding to the civil requirements a further element: that the defendant must

87 Religion is only expressly protected in Queensland, Tasmania and Victoria. 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 (n 19) 125, 143, citing *Jones v Scully* (2002) 120 FCR 243, 271–2 [110]–[113] ('*Scully*'); *Eatock* (n 78) 333 [310].

88 See, eg, Andre Oboler, *Islamophobia on the Internet: The Growth of Online Hate Targeting Muslims* (Report No IR13-7, November 2013); Mariam Veiszadeh, 'Muslim Women Scared to Go Outdoors in Climate of Hate', *The Sydney Morning Herald* (Sydney, 11 October 2014) <www.smh.com.au/comment/muslim-women-scared-to-go-outdoors-in-climate-of-hate-20141009-113p5j>, archived at <https://perma.cc/AFC3-GGRN>.

89 Rees, Rice and Allen (n 67) 627. The relevant federal and state bodies include AHRC, ACT Human Rights Commission, Anti-Discrimination Board of NSW, Anti-Discrimination Commission Queensland and Equal Opportunity Tasmania. South Australia employs a tort action model 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 *Law and Society Review* 631, 641–2.

90 Note that the AHRC does not provide any assistance with bringing a case: 'The Australian Human Rights Commission's Complaint Process: For Complaints about Sex, Race, Disability and Age Discrimination', *Australian Human Rights Commission* (Web Page, 2016) <www.humanrights.gov.au/australian-human-rights-commission-s-complaint-process-complaints-about-sex-race-disability-and-age>, archived at <https://perma.cc/3GLA-DUME>.

91 *Australian Human Rights Commission Act 1986* (Cth) s 46PO.

92 *Discrimination Act 1991* (ACT) s 67A; *Anti-Discrimination Act 1977* (NSW) s 20D; *Anti-Discrimination Act 1991* (Qld) s 131A; *Racial Vilification Act 1996* (SA) s 4; *Racial and Reli-*

threaten physical harm to the person or property of the target person or group, or incite others to threaten harm of that kind.[93] Unlike the civil wrong, there are no statutory defences or exceptions. Again, the Victorian legislation expressly refers to the internet[94] and, unlike the aforementioned jurisdictions, it extends to situations where the offender 'intentionally engages[s] in conduct … likely to incite serious contempt … revulsion or severe ridicule'.[95] 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,[96] as well as corresponding strict liability offences with statutory defences.[97] Apart from one recent case in Queensland,[98] 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 the internet.[99]

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,

gious Tolerance Act 2001 (Vic) s 24; *Criminal Code Act Compilation Act 1913* (WA) ss 77–80D.

93 *Discrimination Act 1991* (ACT) s 67A(1); *Anti-Discrimination Act 1977* (NSW) s 20D; *Anti-Discrimination Act 1991* (Qld) s 131A(1); *Racial Vilification Act 1996* (SA) s 4. NSW, Queensland and South Australia require consent to prosecute from the Attorney-General, a Crown Law Officer, and the Director of Public Prosecutions respectively: see *Anti-Discrimination Act 1977* (NSW) s 20D(2); *Anti-Discrimination Act 1991* (Qld) s 131A(2); *Racial Vilification Act 1996* (SA) s 5.

94 *Racial and Religious Tolerance Act 2001* (Vic) s 24.

95 Ibid s 24(2).

96 *Criminal Code Act Compilation Act 1913* (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.

97 Ibid ss 78, 80, 80B, 80D, 80G.

98 Queensland teenager Abdel Kader Russell-Boumzar pleaded guilty to a number of offences 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* (Sydney, 14 September 2015) <www.abc.net.au/news/2015-09-14/teen-abdel-kader-russell-boumzar-convicted-brisbane-over-abuse/6775454>.

99 *O'Connell v Western Australia* [2012] WASCA 96; Standing Committee on Law and Justice, *Racial Vilification Law in New South Wales* (Report No 50, 3 December 2013) 19 [2.85].

religion, nationality, national or ethnic origin or political opinion'.[100] 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.

B *Criminal Law: Application to the Internet*

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 *Criminal Code Act 1995* (Cth) ('*Commonwealth Criminal Code*').[101] This makes it an offence to use a carriage service[102] 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,[103] meaning if the offender is an Australian citizen, they can be prosecuted even if the conduct occurred wholly outside of Australia.[104] The Explanatory Memorandum for the amending Act[105] 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'.[106] Although there is no reported case law specifically concerning racially motivated conduct,[107] this section has

100 *Criminal Code Act 1995* (Cth) ss 80.2A–80.2B ('*Commonwealth Criminal Code*').

101 A number of other provisions in pt 10.6 (Telecommunications Services) of the Code 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).

102 The *Commonwealth Criminal Code* (n 100) Dictionary (definition of 'carriage services') provides that 'carriage service' has the same meaning as in the *Telecommunications Act 1997* (Cth) s 7 (definition of 'carriage service'): that is, 'a service for carrying communications by means of guided and/or unguided electromagnetic energy'. The internet, then, is clearly a 'carriage service'.

103 This extended jurisdiction applies to all pt 10.6 offences: ibid s 475.2.

104 Ibid s 15.1(1)(c)(i).

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

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

107 In *Starkey v DPP (Cth)* [2013] QDC 124, [5], [7], [51], 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: at [5], [7], [18]–[26].

been employed extensively to deal with harmful conduct online, with 308 successful prosecutions between its introduction in 2005 and 2014.[108]

Significantly, it has been directly and successfully applied in recent years to online conduct of a racially or religiously vilifying nature.[109] 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 n*****', and referenced a desire to 'bash your tall black ass'.[110] The defendant ultimately pleaded guilty and received a conditional release order and A$250 fine.[111] In 2016, a NSW chiropractor was convicted under the section for abusing Indigenous NT former Senator Nova Peris on Facebook, calling her a 'black c***' and demanding that she '[g]o back to the bush and suck on witchity [sic] grubs and yams'.[112]

While these unreported cases suggest that s 474.17 of the *Commonwealth Criminal Code* provides an avenue for redress for online vilification, the conduct in question must reach a high threshold of seriousness. In particular, the Crown must prove that a person used a 'carriage service' and used it in a way that reasonable persons would regard as being, 'in all the circumstances,

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

109 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', *The Sydney Morning Herald* (Sydney, 23 December 2005) <www.smh.com.au/news/national/police-track-text-message-senders/2005/12/22/1135032135717.html>, archived at <https://perma.cc/Q9SP-KARY>.

110 See 'Man Charged over Racist Tweets to Nic Naitanui', *The Age* (Melbourne, 21 March 2014) <www.theage.com.au/afl/afl-news/man-charged-over-racist-tweets-to-nic-naitanui-20140321-hvlb8.html>, archived at <https://perma.cc/MDR7-N8HJ>.

111 Prosecution Notice, *Nguyen* (Magistrates Court of Western Australia, 1239971–2, Magistrate Heaney, 22 July 2014) 3.

112 'Man Who Abused Nova Peris on Facebook Gets Eight-Month Suspended Sentence', *ABC News* (Sydney, 5 July 2016) <www.abc.net.au/news/2016-07-05/man-who-abused-nova-peris-on-facebook-gets-suspended-sentence/7568912>, archived at <https://perma.cc/5PDE-ECVG>; 'Man Handed 8-Month Suspended Sentence for Racist Attack on Nova Peris', *NT News* (Darwin, 5 July 2016) <www.ntnews.com.au/news/northern-territory/man-handed-8month-suspended-sentence-for-racist-attack-on-nova-peris/news-story/d512f15253185c795a7eda4aa48c9582>.

menacing, harassing or offensive'.[113] In *Monis v The Queen*, which dealt with the similarly worded s 471.12,[114] 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'.[115] 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';[116] (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 the attacks.[117] 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 be accepted as offensive but did not meet the threshold for criminal sanction.[118]

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

113 *Commonwealth Criminal Code* (n 100) s 474.17. Section 473.4 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; and (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)'.

114 Ibid. Section 471.12 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(1).

115 *Monis v The Queen* (2013) 249 CLR 92, 210 [336] (Crennan, Kiefel and Bell JJ) ('*Monis (HCA)*'). In *Brown v DPP (Cth)* (2016) 315 FLR 461, 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 (HCA)* when construing s 474.17, given the identical wording of s 474.12 and s 474.17.

116 *Monis v The Queen* (2016) 215 A Crim R 64, 77 [44] (Bathurst CJ).

117 *Monis (HCA)* (n 115) 202–3 [310] (Crennan, Kiefel and Bell JJ).

118 *Starkey* (n 107) [51], [54].

this issue in *Monis*, but the point was not raised in argument, and his Honour expressed no concluded view on the matter.[119]

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 upon the emotional well-being of an addressee'[120] by arousing significant anger, resentment, outrage, disgust or hatred in the mind of a reasonable person.[121]

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,[122] as well as less serious threat offences, which vary by the level of threatened harm required to engage the offence.[123] The requirement of an imminent threat of harm largely rules out the applicability of assault offences to cyber-racism,[124] however, state-based stalking and harassment offences could be

119 *Monis (HCA)* (n 115) 123–4 [45].

120 Ibid 202–3 [310] (Crennan, Kiefel and Bell JJ).

121 Ibid 124 [47] (French CJ), 157 [159] (Hayne J), 199 [299], 200–1 [303] (Crennan, Kiefel and Bell JJ).

122 *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)(a); *Criminal Code Act 1924* (Tas) s 162; *Crimes Act 1958* (Vic) s 20; *Criminal Code Act Compilation Act 1913* (WA) s 338B(a).

123 See, eg, *Crimes Act 1900* (ACT) s 31; *Crimes Act 1900* (NSW) ss 31, 199; *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 Act 1924* (Tas) s 276; *Crimes Act 1958* (Vic) s 21; *Criminal Code Act Compilation Act 1913* (WA) s 338B(b). 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) ss 1 (definition of 'document'), 308; *Criminal Code Act 1924* (Tas) ss 1 (definition of 'writing'), 162.

124 Gregor Urbas, 'Look Who's Stalking: Cyberstalking, Online Vilification and Child Grooming Offences in Australian Legislation' (2007) 10 *Internet Law Bulletin* 62, 62; Simon Bronitt and Bernadette McSherry, *Principles of Criminal Law* (Lawbook Co, 3rd ed, 2010) 563 [10.20]. See, eg, Queensland, Tasmania and Western Australia, where words and images, whether online or otherwise, are insufficient evidence of a threat: *Criminal Code Act 1899* (Qld) s 245; *Criminal Code Act 1924* (Tas) ss 182(1)–(2); *Criminal Code Act Compilation Act 1913* (WA) s 222; Des Butler, Sally Kift and Marilyn Campbell, 'Cyber Bullying in Schools and the Law: Is There an Effective Means of Addressing the Power Imbalance?' (2009) 16(1) *Murdoch University Electronic Journal of Law* 84, 89–90.

used for the prosecution of cyber-harassment, including racial harassment.[125] 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.[126] A number of jurisdictions also have offences pertaining to the use of a computer system to publish or transmit objectionable material.[127] With some presently irrelevant differences between the legislation, each includes a prohibition relating to material that 'promotes crime or violence, or instructs in matters of crime or violence',[128] 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

125 *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 Act 1983* (NT) s 189; *Criminal Code Act 1899* (Qld) s 359B; *Criminal Law Consolidation Act 1935* (SA) s 19AA; *Criminal Code Act 1924* (Tas) s 192; *Crimes Act 1958* (Vic) s 21A; *Criminal Code Act Compilation Act 1913* (WA) s 338E. The *Crimes Act 1900* (NSW) and *Criminal Code Act Compilation Act 1913* (WA) are the only Acts that make no reference to the internet or electronic or technologically assisted forms of communication.

126 Selma Milovanovic, 'Man Avoids Jail in First Cyber Bullying Case', *The Age* (Melbourne, 9 April 2010) <www.theage.com.au/victoria/man-avoids-jail-in-first-cyber-bullying-case-20100408-rv3v.html>, archived at <https://perma.cc/K3H9-SKLL>. The accused pleaded guilty and received an 18-month community sentence, including 200 hours of unpaid community work: Lauren Wilson, 'Cyber Bully Convicted', *The Australian* (Sydney, 9 April 2010) <www.theaustralian.com.au/news/nation/cyber-bully-convicted/news-story/89bf839ef5a49bade777b76d08bcbfe3>, archived at <https://perma.cc/Y9QK-N32E>.

127 *Classification of Publications, Films and Computer Games Act 1985* (NT) ss 77–8; *Classification (Publications, Films and Computer Games) Act 1995* (SA) ss 75C, 75D(1); *Classification (Publications, Films and Computer Games) (Enforcement) Act 1995* (Vic) s 57; *Classification (Publications, Films and Computer Games) Enforcement Act 1996* (WA) ss 101(1)(a), 102.

128 *Classification of Publications, Films and Computer Games Act 1985* (NT) s 75(c) (definition of 'objectionable material'); *Classification (Publications, Films and Computer Games) Act 1995* (SA) s 75A (definition of 'objectionable matter'); *Classification (Publications, Films and Computer Games) (Enforcement) Act 1995* (Vic) s 3 (definition of 'objectionable publication'); *Classification (Publications, Films and Computer Games) Enforcement Act 1996* (WA) s 99(c) (definition of 'objectionable material'). In the *Classification (Publications, Films and Computer Games) Act 1995* (SA) s 75A (definition of 'objectionable matter'), objectionable matter is defined to include 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 *Broadcasting Services Act 1992* (Cth) ('*BSA*') in Part III(C).

are rarely used to prosecute such conduct.[129] The legislation typically refers to conduct occurring in or near, or within view or hearing from, a public place or school or similar,[130] 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.[131]

C *The* BSA *and Cyber-Bullying Legislation*

1 *Online Content Scheme within Schedules 5 and 7 of the* BSA

Another possible avenue of recourse for cyber-racism is the online content scheme within schs 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 eSafety Commissioner ('the Commissioner'), a separate, independent statutory office located within the ACMA.[132] The scheme imposes obligations upon Internet Service Providers ('ISPs')[133] and content/hosting service providers in relation to certain harmful internet content. When the provisions to regulate internet content

129 See David Brown et al, *Criminal Laws: Materials and Commentary on Criminal Law and Process of New South Wales* (Federation Press, 6th ed, 2015) 541–2, citing Luke McNamara and Julia Quilter, 'Turning the Spotlight on "Offensiveness" as a Basis for Criminal Liability' (2014) 39 *Alternative Law Journal* 36, 37–8.

130 *Crimes Act 1900* (ACT) s 392; *Summary Offences Act 1988* (NSW) ss 4(1), 4A; *Summary Offences Act 1923* (NT) ss 47, 53; *Summary Offences Act 1953* (SA) ss 7, 22; *Police Offences Act 1935* (Tas) s 12; *Summary Offences Act 1966* (Vic) s 17; *Criminal Code Act Compilation Act 1913* (WA) ss 74A(1)–(2).

131 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, 3 August 2012) <www.newsinenglish.no/2012/08/03/lawmakers-react-to-bloggers-release/>, archived at <https://perma.cc/8G7W-8BM8>.

132 *Enhancing Online Safety Act 2015* (Cth) ss 14, 15(1); 'Directory: Children's e-Safety Commissioner', *Australian Government* (Web Page, 10 July 2017) <www.directory.gov.au/portfolios/communications-and-arts/australian-communications-and-media-authority/childrens-e-safety-commissioner>, archived at <https://perma.cc/33E6-2E9S>.

133 This is defined in the *BSA* (n 128) sch 5 cl 8(1) as someone who 'supplies, or proposes 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: An Analysis of Codes of Conduct Applying to Online Activity in Australia* (Final Report, March 2012) 38.

were first introduced in 1999,[134] 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.[135] However, it would seem that the scheme has some incidental application 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, sch 7 deals with services that host or provide material online in or from Australia, collectively deemed as 'designated content/hosting service provider[s]'.[136] Both ISPs and content/hosting service providers have obligations imposed upon them by two key regulatory codes registered with the ACMA (as now administered by the Commissioner).[137]

The scheme targets 'prohibited content' or 'potential prohibited content', meaning content that it has been, or is substantially likely to be, given an RC

134 *Broadcasting Services Amendment (Online Services) Act 1999* (Cth).

135 Explanatory Memorandum, Broadcasting Services Bill 1992 (Cth) 66, 69. See also Peter Coroneos, 'Internet Content Policy and Regulation in Australia' in Brian Fitzgerald et al (eds), *Copyright Law, Digital Content and the Internet in the Asia-Pacific* (Sydney University Press, 2008) 49, 52.

136 Schedule 7 applies to 'hosting service', 'live content service', 'links service' and 'commercial content service' providers, collectively known as 'designated content/hosting service provider[s]': *BSA* (n 128) sch 7 cl 2 (definition of 'designated content/hosting service provider'). Under cl 5(1), for the purposes of sch 7, 'a person does not provide a content service merely because the 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.

137 *BSA* (n 128) sch 5 pt 5, sch 7 pt 4; *Enhancing Online Safety Act 2015* (Cth) ss 3, 15; Communications Alliance Ltd, 'Internet Service Providers Voluntary Code of Practice: For Industry Self-Regulation in the Area of Cyber Security' (Industry Code No C650:2014, 2014); 'Internet Industry Compliance', *Office of the eSafety Commissioner* (Web Page) <www.esafety.gov.au/about-the-office/internet-industry-compliance>, archived at <https://perma.cc/3B72-Q25E>. See also Internet Industry Association, 'Codes for Industry Co-Regulation in Areas of Internet and Mobile Content' (Internet Industry Codes of Practice version 10.4, May 2005) ('2005 Industry Code'); Internet Industry Association, 'Internet Industry Code of Practice' (Content Services Code, 10 July 2008); Australian Law Reform Commission ('ALRC'), *Classification: Content Regulation and Convergent Media* (Final Report No 118, February 2012) 53 [2.29].

or X 18+ rating by the Classification Board.[138] Under sch 7, the Commissioner may investigate a complaint into online content at their discretion.[139] 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.[140] Failure to comply with such a notice is an offence, as well as being a civil penalty provision.[141] Where the content meets this threshold, but is hosted overseas, sch 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.[142]

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 ('ALRC'),[143] the RC category is very broad. Under cls 2, 3 and 4 of the *National Classification Code* publications,[144] films and computer games, respectively, will be given an RC rating if they:

> (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.[145]

138 *BSA* (n 128) sch 7 cls 20–1. Schedule 5 indicates that these terms are used in this schedule as defined in sch 7: at sch 5 cl 3 (definition of 'prohibited content' and 'potential prohibited content'). 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).

139 *BSA* (n 128) sch 7 cl 44.

140 Ibid sch 7 pt 3 divs 3–5.

141 Ibid sch 7 cls 106–7. Furthermore, if the Commissioner 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 action directed towards ensuring … [they do] not contravene the rule … in the future'. Failure to comply with a direction under this provision is both an offence and engages civil penalties: at sch 7 cl 108.

142 Ibid sch 5 cl 40. See also the 2005 Industry Code (n 137) cls 19.2–19.3.

143 See ALRC, *Classification*: *Content Regulation and Convergent Media* (n 137) 51.

144 A 'publication' is defined as 'any written or pictorial matter' that is not a film, computer game or 'an advertisement for a publication, a film or a computer game': *Classification (Publications, Films and Computer Games) Act 1995* (Cth) s 5 (definition of 'publication').

145 *National Classification Code* cl 2 item 1.

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 '[t]his means that material relating to drug use, shoplifting, graffiti or euthanasia could … be classified RC'.[146]

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'.[147]

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.

2 *Cyber-Bullying Legislation*

The Commonwealth has also recently introduced cyber-bullying laws, administered by the aforementioned Commissioner.[148] Under the regime, cyber-bullying material[149] must first be reported to the relevant 'social media

146 ALRC, *Classification*: *Content Regulation and Convergent Media* (n 137) 58.

147 *BSA* (n 128) sch 5 cl 40(1)(a), sch 7 cl 69(1).

148 *Enhancing Online Safety Act 2015* (Cth). The government has proposed legislation that would rename the statutory office to that of the 'eSafety Commissioner', and would 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, 'eSafety Office to Help All Australians Online' (Media Release, 9 February 2017) <www.mitchfifield.com/Media/MediaReleases/tabid/70/articleType/ArticleView/articleId/1318/eSafety-office-to-help-all-Australians-online.aspx>, archived at <https://perma.cc/CFB5-UZWB>.

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

service'.[150] For smaller Tier 1 services,[151] 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 Tier 2 services — including Facebook, Google+, Instagram and YouTube[152] — 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 Circuit Court.[153] 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.[154] If they do not comply, the Commissioner can issue a formal warning or obtain an injunction.[155]

Whilst the scheme only applies to seriously harassing, threatening, intimidating or humiliating content targeted at Australian minors, it could encompass content that is racist in nature. However, the legislation deals with material that is directed at a *particular* Australian child.[156] It is therefore unable to account for cyber-racist material directed towards a certain racial

150 'Social media service' is defined very widely under the legislation, and could include social networking platforms, blogging sites and apps, messaging apps which allow content to be included with messages, and video-sharing sites and platforms: see ibid s 9; Office of the Children's eSafety Commissioner, 'Cyberbullying Complaints Handing' (Information Guide Version 3, May 2017) 3–4 <www.esafety.gov.au/complaints-and-reporting/cyberbullying-complaints/complaint-resolution-process>, archived at <https://perma.cc/2533-R9QN>.

151 Any social media service may apply to be a Tier 1 social media service. The application must be approved by the Commissioner if the service complies with basic online safety requirements as prescribed by the legislation and the service is not a Tier 2 social media service: *Enhancing Online Safety Act 2015* (Cth) s 23. The Office of the eSafety Commissioner's website indicates that airG, Ask.fm, Flickr, Snapchat, Twitter, Yahoo!7 Answers and Yahoo!7 Groups are Tier 1 services: 'Social Media Partners: How Our Partners Support the Aims of the Office', *Office of the eSafety Commissioner* (Web Page) <www.esafety.gov.au/social-media-regulation/social-media-partners>, archived at <https://perma.cc/TF3F-S57M>.

152 'Social Media Partners' (n 151). The Minister may declare a social media service to be a Tier 2 social media service on the 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 'the number of accounts … held by end-users who are Australian children', or on request of the social media service provider: *Enhancing Online Safety Act 2015* (Cth) ss 30–1.

153 *Enhancing Online Safety Act 2015* (Cth) ss 35–6, 46–8.

154 Ibid s 42.

155 Ibid ss 43–4, 48(1)(b).

156 Ibid s 5.

group, of which the child is a member, as would be the case for much of the racist material being posted online. Importantly, the scheme recognises that the processes of the social media service itself should act as the first port of call, but provides a backstop against inaction with the ability to enforce penalties against services and perpetrators through the court system. We return to these elements of the scheme later in this article.

D *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 is 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.[157]

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 [Optus's] services or for illegal or unlawful purposes', including use of the service to 'defame, harass or abuse anyone'.[158] They reserve the right to 'block access to, remove, or refuse to post any content' determined to be 'offensive, indecent, unlawful or otherwise inappropriate' regardless of whether it is actually unlawful.[159] 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

157 Citron and Norton (n 4) 1440–2.

158 'Fair Go Policy', *Optus* (Web Page, 22 August 2016) 2 <www.optus.com.au/content/dam/optus/appendix/Appendix%20S/AppS.doc>, archived at <https://perma.cc/V8WA-96WH>.

159 Ibid 3.

complex when we look at major content-hosting platforms based overseas, and the way in which their terms of service operate and interact with the Australian 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,[160] and are not prevented from restricting hate speech in the same way the First Amendment precludes government regulation.[161] 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.[162]

Facebook provides an illustrative example of the efficacy of platform terms of service for dealing with cyber-racist content. In addition to having 13 million active users in Australia,[163] Facebook has been identified as a major site for the proliferation of cyber-racism.[164] 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 'Statement of Rights and Responsibilities' ('Facebook Statement').[165] Under s 3 ('Safety') of the Facebook Statement, 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; or
- use[166] Facebook to do anything unlawful, misleading, malicious or discriminatory.[167]

160 These include Facebook, Instagram (which is owned by Facebook and has over 500 million monthly active users), and YouTube, which is owned by Google.

161 Citron and Norton (n 4) 1439.

162 Oboler and Connelly (n 24) 118.

163 Alex Heber, 'These Incredible Stats Show Exactly How Huge Facebook Is in Australia', *Business Insider Australia* (Sydney, 8 April 2015) <www.businessinsider.com.au/these-incredible-stats-show-exactly-how-huge-facebook-is-in-australia-2015-4>.

164 See Dunn, Paradies and Atie (n 42) 5, 10.

165 'Statement of Rights and Responsibilities', *Facebook* (Web Page, 30 January 2015) <www.facebook.com/legal/terms>, archived at <https://perma.cc/FH8A-G7CV>.

166 The Facebook Statement defines 'use' as 'use, run, copy, publicly perform or display, distribute, modify, translate, and create derivative works of': ibid s 17(7).

Facebook's 'Community Standards'[168] 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 attacks people directly based on, amongst other things, their race, ethnicity, national origin or religious affiliation.[169]

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.[170] 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. Facebook's newsroom indicates that most cases are reviewed within 72 hours, with more serious complaints being prioritised.[171]

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 which many people would deem offensive or constituting 'hate speech'.[172] A pertinent Australian example is the 'Aboriginal Memes' page, which sprung up on Facebook in 2012 denigrating Indigenous Australians.[173] Facebook acknowledged that the page was 'incredibly distasteful' but initially refused to

167 Ibid s 3.

168 'Community Standards', *Facebook* (Web Page, 2017) <www.facebook.com/communitystandards>, archived at <https://perma.cc/427W-NUFR>.

169 Ibid.

170 'How to Report Things', *Facebook* (Web Page, 2017) <www.facebook.com/help/181495968648557>, archived at <https://perma.cc/MYC6-GGRP>.

171 'What Happens after You Click "Report"', *Facebook: Newsroom* (Web Page, 19 June 2012) <https://newsroom.fb.com/news/2012/06/what-happens-after-you-click-report/>, archived at <https://perma.cc/DG2S-SUN5>.

172 See, eg, Andre Oboler, *Aboriginal Memes and Online Hate* (Report No IR12-2, October 2012); Andre Oboler, *Recognizing Hate Speech: Antisemitism on Facebook* (Report No IR13-1, 16 March 2013); Oboler, *Islamophobia on the Internet* (n 88) 2.

173 The Race Discrimination Commissioner, Helen Szoke, at the time described the content of the Facebook page as potentially being in breach of the *RDA*: Emma Sykes, 'Racist Facebook Page Deactivated after Outcry', *ABC News* (Sydney, 9 August 2012) <www.abc.net.au/local/stories/2012/08/08/3563446.htm>, archived at <https://perma.cc/L9TD-G9L5>.

remove it on the grounds that it did not breach its terms of service.[174] The page was then briefly taken down, but re-emerged shortly afterwards with a 'Controversial Humour' tag. Further outcry saw an online petition for its removal gather over 15,000 signatures, and an investigation was commenced by the ACMA.[175] It was finally removed by Facebook, apparently in response to this public pressure.[176] Even then, copycat pages continued to spring up, demanding ongoing intervention from Facebook.[177] 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.[178]

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'.[179] In early 2016, Facebook launched its European 'Initiative for

174 Online Hate Prevention Institute, 'Discussions with Facebook over Aboriginal Memes' (Press Release, 15 August 2012) <http://ohpi.org.au/press-release-discussions-with-facebook-over-aboriginal-memes/>, archived at <https://perma.cc/U5ZW-HCUV>.

175 Sykes (n 173).

176 See ibid.

177 Oboler, *Aboriginal Memes and Online Hate* (n 172) 22–3. See also Rod Chester, 'Facebook Shuts Vile Aboriginal Memes Page, Despite Earlier Claiming It Didn't Constitute "Hate Speech"', *News.com.au* (Australia, 27 January 2014) <www.news.com.au/technology/facebook-shuts-vile-aboriginal-memes-page-despite-earlier-claiming-it-didnt-constitute-hate-speech/story-e6frfrnr-1226811373505earlier-claiming-it-didnt-constitute-hate-speech/story-e6frfrnr-1226811373505>.

178 Oboler and Connelly (n 24) 119. Government request statistics released by Facebook indicate that Facebook 'restricted access in Australia to a number of pieces of content reported under local anti-discrimination laws' between July 2013 and December 2015: see 'Government Requests Report: Australia', *Facebook* (Web Page, 2017) <https://govtrequests.facebook.com/country/Australia/2015-H1/>, archived at <https://perma.cc/PQ6E-RNFG>.

179 Victor Luckerson, 'Facebook, Google Agree to Curb Hate Speech in Germany', *Time* (New York City, 16 December 2015) <http://time.com/4150296/facebook-google-hate-speech-germany/>, archived at <https://perma.cc/AZE3-T7DS>. However, Luckerson states that it is unclear how this will 'affect comments made outside Germany that can still be read by Ger-

Civil Courage Online' and pledged over US$1 million towards stopping extremism on social media.[180] Any effort to better regulate cyber-racism must include and recognise the important role played by these intermediaries, in conjunction with jurisdictional legal frameworks.

E *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 Convention on Cybercrime, concerning the Criminalisation of Acts of a Racist and Xenophobic Nature Committed through Computer Systems*.[181] It requires state parties to criminalise 'making available' or distributing 'racist and xenophobic material … through a computer system' within their domestic laws.[182] Although Australia is party to the *Convention on Cybercrime*,[183] it declined to sign or ratify the Additional Protocol.[184] Outside the governmental sphere, the American-based Anti-Defamation League ('ADL') has released 'Best Practices for Responding to Cyberhate', following consultation between internet providers, civil society organisations, academics, and legal representatives.[185] These soft standards

man 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. In June 2017, the German government passed legislation enabling fines of up to €50 million for social media companies which fail to remove obviously hateful content within 24 hours of being notified, and requiring biannual reports from companies detailing complaints received about hateful content: 'Germany Approves Plan to Fine Social Media Firms up to €50m', *The Guardian* (London, 30 June 2017) <www.theguardian.com/media/2017/jun/30/germany-approves-plans-to-fine-social-media-firms-up-to-50m>, archived at <https://perma.cc/4R7J-HDXT>.

180 Melissa Chan, 'Facebook Launches Initiative to Stop Extremist Posts in Europe', *Time* (New York City, 19 January 2016) <http://time.com/4184559/facebook-initiative-extremism/>, archived at <https://perma.cc/5286-5HP2>.

181 Opened for signature 28 January 2003, CETS No 189 (entered into force 1 March 2006) art 3.

182 Ibid.

183 Opened for signature 23 November 2001, CETS No 185 (entered into force 1 July 2004).

184 AHRC, 'Human Rights in Cyberspace' (n 21) 24.

185 ADL, 'ADL Releases "Best Practices" for Challenging Cyberhate' (Press Release, 23 September 2014) <www.adl.org/press-center/press-releases/discrimination-racism-bigotry/adl-releases-best-practices-challenging-cyberhate.html>, archived at <https://perma.cc/BND9-MLSK>.

provide 'important guideposts' for both intermediaries and the internet community at large.[186]

F *Conclusion*

There is an extensive, if incomplete, network of laws and standards that can be applied to deal with racism on the internet in Australia. 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.[187] In the following section we build on this overview of the regulatory terrain by analysing the relative merits of racial vilification law, criminal law and intermediary codes of conduct 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.

IV Is There a Gap in Regulation?

A *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 line 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 of racial speech. Taking s 18C of the *RDA* as an example, material in the public domain is captured where it is 'offen[sive], insult[ing], humiliat[ing] or intimidat[ing]' on the basis of 'race, colour or national or ethnic origin', as measured objectively from the perspective of a hypothetical reasonable person in the position of the applicant or the applicant's victim group.[188] A strength of this rule is that it does not prohibit generic offence or

186 Ibid.

187 Gelber and McNamara, 'The Effects of Civil Hate Speech Laws' (n 89) 636.

188 AHRC, 'Cyber Racism and Community Resilience Project' (n 74) 6.

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.[189] 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.[190]

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.[191] 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.[192] Yet, 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.[193] Eastman suggests that, 'if asked, the Federal Court may find that a Sikh, Muslim or member of another minority religious communit[y] has an "ethnic origin" for the purpose[s] of the *RDA*'.[194]

Undoubtedly, the process of conciliation inherent in racial vilification laws can be advantageous, allowing for harmful conduct to be dealt with quickly

189 *Creek* (n 78) 356–7 [16].

190 See Wertheim (n 5) 96.

191 In *Eatock* (n 78) 334 [314], the support of the Federal Court for Lord Fraser's approach in *Mandla v Dowell Lee* [1983] 2 AC 548, 562 also suggests that 'ethnic group' may be given a contemporary meaning that is 'appreciably wider than the strictly racial or biological'. 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 and Multicultural Affairs* (1998) 90 FCR 202, 209–11.

192 Eastman (n 19) 146. See, eg, *Scully* (n 87) 272 [113], where ethnic origin under Australian law has been interpreted to include Jews. Ethnic origin under English law has been interpreted to include Sikhs and Roma: *Singh v Rowntree MacKintosh Ltd* [1979] ICR 554; *Commission for Racial Equality v Dutton* [1989] 1 QB 783.

193 Thornton and Luker (n 3) 79–80.

194 Eastman (n 19) 145. Although the Explanatory Memorandum, 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: at 143. 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: at 147. See, eg, *Scully* (n 87) 271–2 [110]–[113]; *Eatock* (n 78) 333 [310].

and informally without resort to the court system. Victims of racist conduct often are not looking for the perpetrator to face heavy penalties, but simply seek a genuine apology acknowledging the harm.[195] 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:

- 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 website owner. Within a few days the website had been disabled by the ISP on account of it breaching their 'Acceptable 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.[196]

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.[197] A small proportion of cases do proceed to a public hearing, creating important precedents that help set community expectations. However, less than 2% of matters under civil vilification laws are resolved in this public manner.[198]

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

195 See Gelber and McNamara, 'The Effects of Civil Hate Speech Laws' (n 89) 647.

196 Conciliation data provided to us by the AHRC indicates that conciliations have occurred with respondent websites, ISPs and social media platforms: see 'Conciliation Register', *Australian Human Rights Commission* (Web Page) <www.humanrights.gov.au/complaints/conciliation-register>, archived at <https://perma.cc/7MHN-DFDP>; 'Race Discrimination: Site Navigation', *Australian Human Rights Commission* (Web Page) <www.humanrights.gov.au/site-navigation>, archived at <https://perma.cc/8SYG-YYTH>.

197 McNamara, *Regulating Racism* (n 3) 310.

198 Gelber and McNamara, 'The Effects of Civil Hate Speech Laws' (n 89) 643. Gelber and McNamara also discuss examples of cases that have resolved in this manner: at 646. See also 'Racial Discrimination Complaints', *Australian Human Rights Commission* (Web Page, 7 November 2016) <www.humanrights.gov.au/news/stories/racial-discrimination-complaints>, archived at <https://perma.cc/FJ22-KZGG>, where, in 2015–16, '[o]nly 3% of complaints finalised by the AHRC were lodged in court'.

in Australia,[199] 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.[200] 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.[201]

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),[202] this depends on proof that this (in)action was connected with the race of the complainant. In *Silberberg v Builders Collective of Australia Inc,* 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.[203] Subsequently in *Clarke v Nationwide News Pty Ltd*, 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.[204] This raises the question of whether the necessary causal 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

199 See *Gutnick* (n 76).

200 AHRC, 'Human Rights in Cyberspace' (n 21) 20.

201 Ibid.

202 *Silberberg v Builders Collective of Australia Inc* (2007) 164 FCR 475, 485 [31]–[34].

203 Ibid 486 [35]. 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: Jonathon Hunyor, 'Cyber-Racism: Can the RDA Prevent It?' (2008) 46(4) *Law Society Journal* (online) 35. In relation to *Silberberg*, Hunyor 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.

204 (2012) 201 FCR 389, 412 [110].

allowing an employer company to be made liable for the actions of an employee (eg a content moderator).[205]

Critically, there is evidence that this model places a heavy burden on the complainant 'to initiate and pursue enforcement proceedings'.[206] 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.[207] 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.[208] 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.[209] This is especially problematic in the online context, given the ease with which harmful material can proliferate.

B *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 attention of the police.[210] Complaints can be

205 Tim Soutphommasane, 'In Defence of Racial Tolerance' (Speech, Australia Asia Education Engagement Symposium, 1 April 2014) <www.humanrights.gov.au/news/speeches/defence-racial-tolerance>, archived at <https://perma.cc/6PQT-96EX>. This still requires the relevant causal connection to be established under s 18C of the *RDA* 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 (n 203) 31.

206 McNamara, *Regulating Racism* (n 3) 310.

207 Gelber and McNamara, 'The Effects of Civil Hate Speech Laws' (n 89) 637.

208 Ibid 646.

209 Ibid 643–4. This may not be the case in all jurisdictions. For example, in 2014–15, the average time it took the AHRC to finalise a complaint was roughly 3.7 months: AHRC, *Annual Report 2014–2015* (n 43) 72.

210 As demonstrated by the prosecution of a woman on a NSW train for offensive behaviour. Her conduct was filmed by a third party on a smart phone and posted online where it was widely condemned as racist: Lucy McNally, 'Karen Bailey Gets Good Behaviour Bond, Avoids Rec-

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,[211] the Commonwealth telecommunications offences appear to offer a satisfactory remedy for an increasing number of complainants (at the less serious end of the spectrum, the utility of offensive language provisions in the online environment remains untested).

Putting these advantages 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 reproduction 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'.[212] 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.[213]

orded Conviction after Racist Train Rant', *ABC News* (Sydney, 1 August 2014) <www.abc.net.au/news/2014-07-31/karen-bailey-avoids-conviction-after-racist-train-rant/5638192>, archived at <https://perma.cc/K4G7-YZHS>. Racially abusive comments on Facebook about former Senator Nova Peris were also the subject of widespread media coverage and a swift public petition asking police to investigate the incident: see, eg, Alina Tooley, 'NSW Police to Investigate Chris Nelson for Racial Vilification of Nova Peris', *Change.org* (Web Page, 2017) <www.change.org/p/nsw-police-nsw-police-to-investigate-chris-nelson-for-racial-vilification-of-nova-peris>, archived at <https://perma.cc/QFW9-2NXP>.

211 See *Monis (HCA)* (n 115) 202–3 [310]. See also Part III(B)(1) for a discussion of how the term 'offensive' has been interpreted in relation to the telecommunications services section of the *RDA*.

212 See McNamara, *Regulating Racism* (n 3) 247.

213 The exceptions to this are the criminal racial vilification provisions: *Criminal Code 2002* (ACT) s 750; *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 Act Compilation Act 1913* (WA) ss 77–80D. 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).

C *Intermediary Terms of Service and Codes of Conduct*

By way of contrast, terms of service and codes of conduct adopted by intermediaries 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 perpetrator uses a pseudonym or is located overseas.

There continue to be questions about the extent to which online intermediaries could be directly liable for failing to remove racist material located on their platforms under the 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 reviewing and removing the majority of racist content within 24 hours.[214] It seems this agreement was developed as a compromise to avoid the European Commission or individual countries imposing greater liability on the platforms.[215]

D *Conclusion: A Gap*

The internet, by its very nature, presents significant regulatory challenges stemming from the quantity of activity, the ease of anonymity and its border-

[214] European Commission, 'European Commission and IT Companies Announce Code of Conduct on Illegal Online Hate Speech' (Press Release No IP/16/1937, 31 May 2016) <http://europa.eu/rapid/press-release_IP-16-1937_en.htm>, archived at <https://perma.cc/83VV-LDUF>.

[215] 'European Union Agreement with Social Media Platforms on Tackling Hate Speech', *The Online Hate Prevention Institute* (Web Page, 31 May 2016) <http://ohpi.org.au/european-union-agreement-with-social-media-platforms-on-hate-speech/>, archived at <https://perma.cc/JQQ7-C2GP>.

less nature where 'everything goes'.[216] 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 or weakness in the regulation of cyber-racism in Australia. There is no comprehensive process 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 of enforcement that engages intermediaries, perpetrators, bystanders and victims. In the following section we consider measures that might help address this gap.

V Addressing the Gap

A *The Applicability of Australia's Cyber-Bullying Legislation*

A recent 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.[217] Of equal significance is the finding that 80% of survey respondents support laws against racial vilification, and close to 70% support laws against religious vilification.[218] 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.

216 Sarah Rohlfing, 'Hate on the Internet' in Nathan Hall et al (eds), *The Routledge International Handbook on Hate Crime* (Routledge, 2015) 293, 297.

217 Dunn, Paradies and Atie (n 42).

218 Andrew Jakubowicz et al, 'What Do Australian Internet Users Think about Racial Vilification?', *The Conversation* (Melbourne, 17 March 2014) <http://theconversation.com/what-do-australian-internet-users-think-about-racial-vilification-24280>, archived at <https://perma.cc/U2DB-Z4AD>.

Though recommendations for greater control over online content often create additional pressure for criminalisation,[219] criminal law has not been the 'preferred vehicle' of Australian legislatures for 'regulating racial vilification'.[220] 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.[221] The New Zealand government has also enacted comparable but broader reforms through their *Harmful Digital Communications Act*.[222] Importantly, both schemes include mechanisms that engage with end-users (those who post harmful content), and the platforms which host harmful material.

As already noted, 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.[223] The scheme allows for court orders and potential fines on both large social media platforms[224] and on end-users who fail to remove cyber-bullying material.[225] Crucially, the 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.[226] The legislation sets out 'basic online safety requirements for a social

219 See Audrey Guichard, 'Hate Crime in Cyberspace: The Challenges of Substantive Criminal Law' (2009) 18 *Information and Communications Technology Law* 201, 224; Oboler, *Aboriginal Memes and Online Hate* (n 172) 63–4. On the criminalisation of 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); Nicola Henry and Anastasia Powell, 'Sexual Violence in the Digital Age: The Scope and Limits of Criminal Law' (2016) 25 *Social and Legal Studies* 397, 404.

220 McNamara, *Regulating Racism* (n 3) 204, 308.

221 *Enhancing Online Safety Act 2015* (Cth). The federal government is also presently consulting on the efficacy of a civil model to deal with the non-consensual sharing of intimate images, drawing upon aspects of the cyber-bullying regime: Department of Communications and the Arts, 'Civil Penalties Regime for Non-Consensual Sharing of Intimate Images' (Discussion Paper, May 2017).

222 *Harmful Digital Communications Act 2015* (NZ).

223 *Enhancing Online Safety Act 2015* (Cth) s 29.

224 Ibid ss 35–6, 46–8.

225 Ibid ss 42–4, 48.

226 Ibid s 15(1).

media service'.[227] 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.[228] In this way, intermediaries are incentivised to improve their online safety and reporting practices.

The Australian legislation was inspired in part by the New Zealand legislation, which at that time was still being debated.[229] The New Zealand scheme employs an 'Approved Agency'[230] to investigate complaints about harm, defined as 'serious emotional distress',[231] caused to individuals by digital communications.[232] Provided that an affected individual has already brought a claim before the Approved Agency,[233] they may bring a case in the District Court, which has the power to make orders against end-users and online content hosts.[234] Non-compliance with an order without reasonable excuse is

227 Ibid s 21 (emphasis altered).

228 Ibid s 23. As outlined in the legislation, only Tier 2 social media services are subject to punitive enforcement under the Act: at ss 36, 47–8.

229 Explanatory Memorandum, Enhancing Online Safety for Children Bill 2014 and Enhancing Online Safety for Children (Consequential Amendments) Bill 2014 (Cth) 51.

230 *Harmful Digital Communications Act 2015* (NZ) s 7. 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' (Release, 31 May 2016) <www.beehive.govt.nz/release/netsafe-appointed-cyberbullying-role>, archived at <https://perma.cc/QE58-LQXE>; 'Netsafe Starts New Role Dealing with Online Harassment', *New Zealand Law Society* (Web Page, 24 November 2016) <www.lawsociety.org.nz/news-and-communications/latest-news/news/netsafe-starts-new-role-dealing-with-online-harassment>, archived at <https://perma.cc/7YC3-QFJB>.

231 *Harmful Digital Communications Act 2015* (NZ) s 4 (definition of 'harm').

232 Ibid s 8.

233 Claims may also be brought by a parent or guardian, a school on behalf of an affected student with consent, or the police: ibid s 11.

234 Ibid ss 12(1), 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, a serious breach, or a repeated breach of 1 or more communication principles; and … the breach has caused or is likely to cause harm to an individual'. Relevant communication principles for our purposes, as outlined in s 6(1), include Principle 2 ('A digital communication should not be threatening, intimidating, or menacing'), Principle 8 ('A digital communication should not incite or encourage anyone to send a message to an individual for the purpose of causing harm to the individual') and Principle 10 ('A digital communication should not denigrate an individual by reason of his or her colour, race, ethnic or national origins, religion, gender, sexual orientation, or disability').

a criminal offence.[235] 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 the author cannot be contacted or does not respond within 48 hours, the host must remove the content.[236]

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.[237] It also places the onus on the online content host to contact perpetrators, rather than encouraging complainants to report to the intermediary where there are existing reporting channels (eg where it is some kind of social media service). 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 threatening, seriously intimidating, seriously harassing or seriously humiliating' a child[238] — is both less vague than the New Zealand standard and more in line with other Australian provisions regulating speech. The Australian regime also recognises the importance of existing procedures employed by intermediaries to combat harmful speech, and the need to educate and empower internet users to employ those channels. The New Zealand legislation introduces a new broad criminal offence of causing harm by posting digital communication.[239] This is arguably unnecessary in Australia in light of existing Commonwealth telecommunications offences.

For these reasons, the Australian cyber-bullying scheme provides a more promising statutory model for exploring how best to confront the gap in protection from online racism.[240] It presents a new and unparalleled oppor-

235 Ibid s 21.

236 Ibid s 24.

237 See, eg, New Zealand Parliament, *Parliamentary Debates*, 30 June 2015, 4851 (Jacinda Ardern), 4862 (Gareth Hughes).

238 *Enhancing Online Safety Act 2015* (Cth) s 5.

239 *Harmful Digital Communications Act 2015* (NZ) s 22.

240 This potential was raised in consultations prior to the enactment of the legislation: Online Hate Prevention Institute, Submission to the Department of Communications, *Enhancing Online Safety for Children* (7 March 2014) <www.communications.gov.au/sites/g/files/net301/f/submissions/Online_Hate_Prevention_Institute.pdf>, archived at <https://perma.cc/LDY7-H78Y>.

tunity 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 for strengthening the tools already available to remedy the harm of cyber-racism, illustrated by a brief hypothetical scenario on how these elements might come together to operate in practice. While we present this as a new model, our primary intention is to explore the merits of these individual elements rather than convince readers of the necessity of establishing a bespoke civil enforcement regime for cyber-racism, which we see as an open question.

B *Elements of a Civil Penalties Approach*

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 *Commonwealth Criminal Code*, and s 18C of the *RDA*.[241]

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[242] that has been in place for over 20 years.[243] As we noted above, both internet users[244] and the general community[245] have recently

241 The scope of this legislation is confined to children as users of the internet: *Enhancing Online Safety Act 2015* (Cth) s 5. Such a limitation would be unduly restrictive in the context of cyber-racism, which can inflict harm on persons of all ages.

242 Tim Soutphommasane, 'A Brave Act' (Conference Paper, 40 Years of the *Racial Discrimination Act 1975* (Cth) Conference, 19–20 February 2015) 7, 9 <www.humanrights.gov.au/our-work/race-discrimination/publications/perspectives-racial-discrimination-act-papers-40-years>, archived at <https://perma.cc/RCE9-4BUZ>.

243 *Racial Hatred Act 1995* (Cth).

244 Jakubowicz et al (n 218).

245 'Overwhelming Majority Reject Change to Racial Vilification Law' (n 30). 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

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'.[246] That being said, the current definition of racial vilification under the *RDA* is not 'the last word on the matter'.[247] As the Parliamentary Inquiry into Freedom of Speech shows, the extent to which s 18C strikes the right balance between the fundamental values of freedom of expression and freedom from racism is a matter of ongoing debate.[248] This is well exemplified by the failure of the Parliamentary Joint Committee on Human Rights to reach agreement on recommended changes to the wording of s 18C.[249]

Legislative review that carefully calibrates community viewpoints, without falling hostage to narrow political interests at either end of the spectrum, has the potential to help ensure that legal standards are in keeping with public attitudes.[250] However, no law 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). Irrespective of whether Parliament decides to change the s 18C threshold, it

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.

246 Luke McNamara and Katharine Gelber, 'The Impact of Section 18C and Other Civil Anti-Vilification Laws in Australia' (Conference Paper, 40 Years of the *Racial Discrimination Act 1975* (Cth) Conference, 19–20 February 2015) 167 <www.humanrights.gov.au/our-work/race-discrimination/publications/perspectives-racial-discrimination-act-papers-40-years>, archived at <https://perma.cc/E3VE-XVTY>.

247 Geoffrey Brahm Levey, 'Why the Campaign to Reform the *Racial Discrimination Act* Failed' (Conference Paper, 40 Years of the *Racial Discrimination Act 1975* (Cth) Conference, 19–20 February 2015) 98 <www.humanrights.gov.au/our-work/race-discrimination/publications/perspectives-racial-discrimination-act-papers-40-years>, archived at <https://perma.cc/E3VE-XVTY>.

248 Parliamentary Joint Committee on Human Rights (n 29) 5.

249 Ibid ix. Recommendation 3 of the Committee outlines six 'proposals that had the support of at least one member of the committee'. These include: 'no change to sections 18C or 18D'; codifying the existing judicial interpretations of s 18C as 'profound and serious effects not to be likened to mere slights'; replacing the words 'offend', 'insult' and 'humiliate' in s 18C with 'harass'; 'changing the objective test from "reasonable member of the relevant group" to "the reasonable member of the Australian community"'; 'amending section 18D to also include a "truth" defence'; and further investigating 'criminal provisions on incitement to racially motivated violence': at ix–x.

250 Levey (n 247) 98–9.

provides the most appropriate and widely accepted framework for setting comparable online standards.

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. These comments 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.[251]

Although this threshold allows for a fairly broad interpretation of racist speech, one that is consistent with our earlier definition of cyber-racism, 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,[252] the 'racialisation of religious belief' has been criticised by Thornton and Luker for normalising some religions, such as Christianity, at the expense of others,

251 McNamara, *Regulating Racism* (n 3) 249.

252 Eastman (n 19) 146. 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': New South Wales, *Parliamentary Debates*, Legislative Council, 4 May 1994, 1827–8 (JP 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 *Anti-Discrimination Act 1977* (NSW): see Eastman (n 19) 144–5. Vilification based on religious belief is more explicitly prohibited in Queensland, Tasmania and Victoria: *Anti-Discrimination Act 1991* (Qld) ss 124A, 131A; *Anti-Discrimination Act 1998* (Tas) s 19; *Racial and Religious Tolerance Act 2001* (Vic) ss 8, 25.

particularly Islam.[253] While attempts to legislate for cyber-racism risk duplicating the definitional 'confusion and ambiguity'[254] that troubles all discrimination and vilification law, such uncertainty, and the controversy it attracts, is inherent in this field of law. It would be counterproductive to see it as an insurmountable obstacle to stronger regulation of cyber-racism.

2 *Utilisation of Existing Intermediary Reporting Mechanisms*

Attempts to tighten the regulation of racial comments online cannot ignore the key role played by intermediaries 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 a scheme of this nature.

This approach has the advantage of placing responsibility upon intermediaries for picking up some of the financial cost of regulating their own platforms. A shortcoming of this element is that it places responsibility on the individual to make an initial complaint, and intermediaries 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.

3 *Pressure on Intermediaries to More Effectively Police Online Conduct, Including Liability for Failure to Respond*

Enacting firm legislation of this sort puts pressure on intermediaries to improve their mechanisms for dealing with racist conduct and enforce existing codes more effectively, so as to avoid civil penalties. The potential of this approach for cyber-racism is evidenced in the apparent success of the cyber-bullying legislation in its first 12 months of operation. Under that

253 Thornton and Luker (n 3) 74–5, 91.

254 Ibid 72.

legislation, larger Tier 2 social media platforms risk a A$21,000 daily fine if they fail to comply with a takedown notice within 48 hours,[255] but heartening-ly, many have responded in less than a day,[256] meaning there is no need to fall back on civil penalties.

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 fine for non-compliance.

By placing responsibility on intermediaries to respond to complaints, including by removing of material, the advantage of this element is that it bypasses 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 that their terms of service clearly address the harmful behaviour. In the cyber-bullying legislation, for example, smaller platforms can be given Tier 1 status, reflecting the intention that the scheme operate in partnership with social media services.[257] 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. The Commissioner may also recommend that the service be declared a Tier 2 service, thereby subjecting

255 *Enhancing Online Safety Act 2015* (Cth) ss 35–6, 46.

256 Australian Communications and Media Authority and Office of the Children's eSafety Commissioner, *Annual Reports 2015–16* (Annual Report, 2016) 122–4. See also Sunanda Creagh, 'Full Response from a Spokesperson for Mitch Fifield', *The Conversation* (Melbourne, 29 August 2016) <http://theconversation.com/full-response-from-a-spokesperson-for-mitch-fifield-64439>, archived at <https://perma.cc/292G-6F9Q>.

257 'About Tier 1 of the Scheme', *Office of the eSafety Commissioner* (Web Page) <www.esafety.gov.au/social-media-regulation/about-tier-1-of-the-scheme>, archived at <https://perma.cc/6FYP-AC8S>.

the service to the enforcement mechanism.[258] However, the efficiency of this process for unresponsive intermediaries is untested.

A further shortcoming of the cyber-bullying enforcement mechanism is that the maximum A$21,000 daily penalty is a very small 'stick' given the high profit margins of major online platforms.[259] A higher penalty would certainly not be unreasonable to deal more effectively with harmful online content.

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.[260] Involving non-victim parties in regulation would help build community capacity to identify and respond to racist behaviour and take the pressure off those who are its targets.

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 the legislative standards can be commenced and conducted', combined with the relatively broad scope of the legislation.[261] By the same token, the cyber-bullying scheme provides a simple

258 Ibid.

259 The most significant example of this is Facebook, which reported profits in excess of US$1 billion in the final quarter of 2015: Deepa Seetharaman, 'Facebook Profit Tops $1 Billion', *Wall Street Journal* (New York City, 27 January 2016) <www.wsj.com/articles/facebook-profit-tops-1-billion-1453929139>.

260 Rivkah Nissim, 'Building Resilience in the Face of Racism: Options for Anti-Racism Strategies' (Discussion Paper, 27 October 2014) 4 <http://apo.org.au/node/41961>, archived at <https://perma.cc/NVP8-5F5Q>.

261 McNamara, *Regulating Racism* (n 3) 311.

online complaint process[262] 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.[263] By minimising the burden on individual victims this kind of 'collective response'[264] goes some way towards recognising that racial vilification constitutes a public, not just an individual, wrong.

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 take-down 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). Having a legal threat overhead 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.

262 See 'Report Cyberbullying', *Office of the eSafety Commissioner* (Web Page) <www.esafety.gov.au/complaints-and-reporting/cyberbullying-complaints/i-want-to-report-cyberbullying>, archived at <https://perma.cc/2XSA-3UWT>.

263 'Cyberbullying FAQs', *Office of the eSafety Commissioner* (Web Page) <www.esafety.gov.au/complaints-and-reporting/cyberbullying-complaints/cyberbullying-complaints-faqs>, archived at <https://perma.cc/D9NJ-NDXG>. For an example of a protracted complaint under the civil human rights system of conciliation, which took nearly five years to be resolved, see *Trad v Jones [No 3]* [2009] NSWADT 318. For the subsequent legal proceedings and the question of costs, see Gelber and McNamara, 'The Effects of Civil Hate Speech Laws' (n 89) 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 *International Journal of Police Science and Management* 388, 388–9.

264 Gelber and McNamara, 'Private Litigation to Address a Public Wrong' (n 3) 333.

6 *Mechanisms to Educate Internet Users*

Despite the benefits of the confidential conciliation process used in civil vilification laws,[265] research suggests that the public continue to have an uneven or limited knowledge about the rules that govern both racial vilification[266] and harmful online speech.[267] This reinforces the need for an educational function to be built into any attempt to strengthen the regulation of cyber-racism.

By encouraging access to online reporting mechanisms, the cyber-bullying approach can help equip users, including many who are young people,[268] 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.[269] Another advantage is that an order made under such a scheme would be available in the public domain to play an educative role about inappropriate online behaviour.[270] If these elements were integrated into any future strategy for tighter regulation of cyber-racism, 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

265 See Gelber and McNamara, 'The Effects of Civil Hate Speech Laws' (n 89) 643.

266 See ibid; AHRC, 'Human Rights in Cyberspace' (n 21) 15.

267 See New Zealand Law Commission, 'Harmful Digital Communications' (n 48) 60–1 [3.43]–[3.52].

268 A survey of social media use in Australia in 2015 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: see Sensis, *How Australian People and Businesses Are Using Social Media* (Report, May 2015) 11, 14. ACMA research into Australian teenagers using the internet found that 72% accessed the internet more than once a day, with the most time spent on Facebook and YouTube amongst the most accessed platforms: 'Aussie Teens Online', *Australian Communications and Media Authority* (Web Page, 1 July 2014) <www.acma.gov.au/theACMA/engage-blogs/engage-blogs/Research-snapshots/Aussie-teens-online>, archived at <https://perma.cc/9DN4-C8CY>.

269 See Nissim (n 260) 3.

270 See, eg, *Australian Cybercrime Online Reporting Network* (Website) <www.acorn.gov.au/>, archived at <https://perma.cc/ZJJ5-6VDW>. The Australian Cybercrime Online Reporting Network ('ACORN') was launched in November 2014 to offer a simple and streamlined way to report cybercrime and ensure reports are referred to the correct agency: Paul Osborne, 'New Online Tool ACORN Allows Australians to Report Cybercrime in Real Time', *The Sydney Morning Herald* (Sydney, 26 November 2014) <www.smh.com.au/digital-life/consumer-security/new-online-tool-acorn-allows-australians-to-report-cybercrime-in-real-time-20141125-11u0v1.html>, archived at <https://perma.cc/ZT48-QADC>.

permit'.[271] 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.

7 *Enhancement of the Ability to Record and Monitor Online Behaviour*

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.[272] 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 intermediaries. This information has the potential to enhance community understanding about appropriate online standards.

C *The Administration of a Civil Penalties Scheme*

This leaves us with 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.[273] 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 inter-

271 Gelber and McNamara, 'The Effects of Civil Hate Speech Laws' (n 89) 649.

272 See, eg, Online Hate Prevention Institute, *Fight against Hate* (Website) <https://fightagainsthate.com/>, archived at <https://perma.cc/B3TH-N44L>.

273 AHRC, *Corporate Plan 2015–2016* (Plan, July 2015) 3, 6.

est',[274] the 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 television.[275] In addition, it is already empowered to undertake enforcement action, including applications to the Federal Court for certain orders and civil penalties.[276] 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 eSafety Commissioner would 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.[277] Recent legislation has formalised the expansion of the office to promote online safety for all Australians, not just primarily children.[278] The elements proposed above thus nicely complement the functions of the Commissioner.

D *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

274 'Introduction to the ACMA', *Australian Communications and Media Authority* (Web Page, 5 September 2016) <www.acma.gov.au/theACMA/About/Corporate/Authority/introduction-to-the-acma>, archived at <https://perma.cc/LN7Y-RSBD>.

275 *BSA* (n 128) sch 7 pt 9.

276 For example, '[i]f the ACMA is satisfied that a person is providing subscription radio broadcasting services … [other] than in accordance with the relevant class licence … [it] may apply to the Federal Court for an order that the person cease providing those services': ibid s 144(1). The ACMA may apply to the Federal Court to enforce various civil penalty orders: at ss 205F–205G.

277 In April 2016, the federal government launched 'eSafetyWomen', an initiative of the Office of the Children's eSafety Commissioner: Mitch Fifield and Michaelia Cash, 'New eSafety Women Website Launched' (Media Release, 28 April 2016) <www.mitchfifield.com/Media/MediaReleases/tabid/70/articleType/ArticleView/articleId/1150/joint-media-release--New-eSafety-Women-website-launched.aspx>, archived at <https://perma.cc/2WS7-G2CU>.

278 See *Enchancing Online Safety Act 2015* (Cth); Explanatory Memorandum, Enhancing Online Safety for Children Amendment Bill 2017 (Cth).

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 more comprehensive regulatory system 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.

VI Conclusion

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 a 'hands-off' approach. Any lack of will to regulate is resoundingly countered by evidence showing support for careful and tailored intervention.

This article has surveyed the regulatory channels for dealing with cyber-racism in Australia. 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. This analysis exposes 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 weakness in protection, we have emphasised a multi-pronged approach that places greater responsibility on industry codes of conduct, reinforced by state intervention. We have explored the contribution that key elements of a civil-penalties-type model could make to an enforceable scheme, once voluntary measures are exhausted — a scheme which deals expressly with online content and targets both perpetrators and third-party hosts.

All legislative reform creates definitional debate. This uncertainty should not be seen as a fundamental obstacle to reform. After all, while some Australians bombarded David Jones with racist speech after their appoint-

ment of Goodes as brand ambassador, there were many others who condemned that conduct, online, in the media and in the broader community. Whatever the challenges associated with regulating online conduct, Australian citizens are willing to speak out against racism, and should be empowered with the best possible tools to do so.