

BUILDING TOLERANCE INTO HATE SPEECH LAWS: STATE AND TERRITORY ANTI-VILIFICATION LEGISLATION REVIEWED AGAINST INTERNATIONAL LAW STANDARDS

NICHOLAS ARONEY* AND PAUL TAYLOR†

United Nations ('UN') monitoring bodies frequently pose questions about Australia's compliance with the hate speech mandates of key UN conventions. Recently, the Human Rights Committee enquired about inconsistencies across Australian state and territory anti-vilification legislation, as raising issues under the International Covenant on Civil and Political Rights ('ICCPR'). This article examines the implications of those inconsistencies, both legal and practical, for Australia's ICCPR compliance. At a time when hate speech is not abating but becoming a common feature of an increasingly fragmented society, this article asks the following questions: Are the settings for anti-vilification legislation at state and territory levels appropriate in the balance achieved between applicable human rights? Can Australian state and territory legislation be better targeted to distinguish between prohibited and preserved free speech? Do Australian state and territory laws conform to the requirements of the ICCPR and other UN instruments? The article concludes that the inconsistencies are problematic and lead to public uncertainty, exacerbated by the unpredictable application by some competent authorities. It proposes legislative solutions that focus less on the scope of prohibition (which is dependent on terminology lacking definitional precision) and more on bringing clarity to the scope of excluded conduct, in conformity with ICCPR demands protecting freedom of expression.

* Professor of Constitutional Law, TC Beirne School of Law, The University of Queensland; Executive Director (Public Law) of the Centre for Public International and Comparative Law ('CPICL'), The University of Queensland.

† Honorary Senior Lecturer, TC Beirne School of Law, The University of Queensland; Fellow of CPICL, The University of Queensland; Adjunct Professor, School of Law, The University of Notre Dame Australia. The authors acknowledge the generosity of the Estate of Douglas Slatter and Elizabeth Chambers in the provision of funding that supported the writing of this article. We also express our thanks to Associate Professor Neville Rochow KC for commenting on an earlier version of this article.

I INTRODUCTION

In June 2019, the United Nations ('UN') Secretary-General, Antonio Guterres, gave renewed urgency to the pursuit of wide-ranging efforts that address the root causes of hate speech, when launching the *United Nations Strategy and Plan of Action on Hate Speech*:

Hate speech is in itself an attack on tolerance, inclusion, diversity and the very essence of our human rights norms and principles. More broadly, it undermines social cohesion, erodes shared values, and can lay the foundation for violence, setting back the cause of peace, stability, sustainable development and the fulfilment of human rights for all.¹

A case can be made that protective measures against 'hate speech' are necessary, but much turns on what is meant by that term and the appropriateness of the measures deployed to meet its harms. As a matter of international law, incitement to racist discrimination, as well as acts of violence or incitement to such acts must be met by criminal law sanctions. Non-criminal prohibitions are needed to meet less extreme harms. A suite of non-legislative measures should also be applied, including those directed at educating and influencing attitudes.

This article considers the different degrees of legislative intervention required by international law, and the key conditions to be met. Its focus is Australian state and territory 'anti-vilification' laws, as hate speech laws tend to be called in the Australian context. Their rationale was lucidly expressed by Katharine Gelber and Luke McNamara: '[i]n a society that aspires to embrace diversity and support the human rights of all, it is not OK to vilify someone (that is, denigrate or defame them) because of who they are, as opposed to something they might have done.'² This principle is attractive, and one might wish that it guided the interactions of all members of Australian society. The question is whether legal regulation is appropriate and, if so, exactly how such laws should be framed.³ Vilifying speech at an appropriate threshold arguably warrants regulation because of its destructive capability. It usually targets the community's most vulnerable, and it can affect an individual's or entire group's ability to participate fully in the ordinary activities that most Australians take for granted. This article therefore assumes the need for some regulation of hate

¹ Antonio Guterres, 'United Nations Strategy and Plan of Action on Hate Speech' (UN Office on Genocide Prevention and the Responsibility to Protect, 18 June 2019) ('Strategy and Plan of Action').

² Katharine Gelber and Luke McNamara, 'Why Australia's Anti-Vilification Laws Matter', *The Conversation* (online, 30 November 2018) <<https://theconversation.com/why-australias-anti-vilification-laws-matter-106615>>. See generally Katharine Gelber and Adrienne Stone (eds), *Hate Speech and Freedom of Speech in Australia* (Federation Press, 2007); Jeremy Waldron, *The Harm in Hate Speech* (Harvard University Press, 2012); Alex Brown, *Hate Speech Law: A Philosophical Examination* (Routledge, 2015).

³ Rex Tauati Ahdar, 'Religious Vilification: Confused Policy, Unsound Principle and Unfortunate Law' (2007) 26(2) *University of Queensland Law Journal* 293.

speech, but its focus is more confined. It asks what is fitting by way of *legislative* response to different forms of hate speech. It enquires whether state and territory statutes achieve targeted intervention against vilifying speech at the appropriate threshold, and by the best available means, with due regard for the rights affected and, if not, whether corrective steps are needed. These are aimed at bringing Australian anti-vilification legislation into closer conformity with the requirements of international law. As a former Special Rapporteur on Freedom of Religion or Belief has observed, ‘failure to act on “real” incitement cases’ and ‘overzealous reactions to innocuous cases’ tends to create ‘a climate of impunity for some and a climate of intimidation for others’.⁴ It is vital, therefore, that an appropriate response to hate speech is combined with a proper respect for freedom of expression.

Australia is bound by the *International Convention on the Elimination of All Forms of Racial Discrimination* (‘ICERD’)⁵ and the *International Covenant on Civil and Political Rights* (‘ICCPR’).⁶ It is periodically required to report and face questioning on the state of the domestic implementation of its obligations before the convention monitoring bodies, the Committee on the Elimination of All Forms of Racial Discrimination (‘CERD Committee’) and the Human Rights Committee. One issue recently raised by the Human Rights Committee in connection with the ICCPR concerns the existence of inconsistencies in anti-vilification laws across different states and territories in Australia.⁷ The Committee asked why this is the case and was also keen to know whether there were plans to introduce federal legislation to reconcile the inconsistencies.⁸

This article addresses these questions by examining the extent of the inconsistencies across state and territory anti-vilification laws, and identifying key areas in which legislators should have closer regard to international standards. It concludes that it is both appropriate and necessary to consider a federal initiative, applying principles for hate speech established by the Human Rights Committee and other UN sources. A companion article addresses the corresponding question in relation to the relevant Commonwealth legislation.⁹

⁴ Heiner Bielefeldt, *Report of the Special Rapporteur on Freedom of Religion or Belief*, UN Doc A/HRC/31/18 (23 December 2015) [63].

⁵ *International Convention on the Elimination of All Forms of Racial Discrimination*, opened for signature 21 December 1965, 660 UNTS 195 (entered into force 4 January 1969) (‘ICERD’).

⁶ *International Covenant on Civil and Political Rights*, opened for signature 16 December 1966, 999 UNTS 171 (entered into force 23 March 1976) (‘ICCPR’).

⁷ Human Rights Committee, *Concluding Observations on the Sixth Periodic Report of Australia*, UN Doc CCPR/C/AUS/CO/6 (16 October–10 November 2017) 19, 20.

⁸ Human Rights Committee, *Summary Record of the 3419th Meeting — Consideration of Reports Submitted by States Parties Under Article 40 of the Covenant* Australia, UN Doc CCPR/C/SR.3419 (24 October 2017) 4 [21].

⁹ Nicholas Aroney and Paul Taylor, ‘The Rights and Wrongs of 18C: An International Perspective on the Racial Hatred Provisions of the Racial Discrimination Act 1975’ (2023) *Australian International Law Journal* (forthcoming) (‘The Rights and Wrongs of 18C’).

There are three main concerns. The first is the potential reach of *civil* law prohibitions at the lower end of the spectrum, closer to the threshold for liability (rather than more extreme hate speech, currently met by *criminal* measures). In particular, the legislation relies on terminology that does not lend itself to precision (such as ‘offend’, ‘hatred’, ‘severe ridicule’ and ‘serious contempt’). The effectiveness of anti-vilification laws depends, for success, on self-censorship, but such uncertainty comes with a cost to personal and democratic freedoms. Those against whom a complaint is allowed to progress face the stigma of a hate speech accusation, as well as the spectre of considerable cost in defending themselves, even if a tribunal ultimately finds that the law has no application.

In this article we argue that this risk is best cured by an objective mechanism that excludes from prohibition conduct for which no justifiable basis exists for its restriction, according to the standards established for freedom of expression in art 19 of the *ICCPR*. Article 19 defines what should be guaranteed and gives States Parties scope to limit the freedom in upholding the rights of victims of hate speech. Article 20(2) mandates a particular form of restriction on the freedom that manifestly accords with the requirements of art 19. Legislation implementing the requirements of art 20(2) would, therefore, answer the Human Rights Committee’s concerns in a manner that complies with art 19. Following this general lead, our proposed solution focuses not so much on the ambit of civil prohibitions, but rather aims to achieve greater precision in what is excluded.

The second concern is that the material divergence across state and territory civil anti-vilification legislation (in such matters as the applicable grounds for the prohibitions, the scope of prohibited conduct, and the conceptual basis for liability) can make it extremely difficult for the public, even those conscientiously contributing to matters of public debate, to know with any certainty whether their speech is unlawful. The third problem is that intervention of competent human rights commissions or other authorities, charged at state and territory level with facilitating the conciliation of complaints and taking other action, can be powerfully influential. They are generally required to dismiss a complaint if it is obvious that the law has no application in the circumstances. If complaints are not dismissed when they should be, the likely result is to place, in the public mind, the belief that the impugned conduct is unlawful, when it is not. If the legislation is not clear and predictable, it is hard for authorities to act confidently either way, without bringing themselves, and the legislation they administer, into disrepute.

The article is structured as follows. It begins in Part II by contrasting the mechanisms at international law and under Australian state and territory anti-vilification legislation for concurrently upholding freedom of expression and protecting against hate speech. Part III takes up the issue raised by the Human Rights Committee at Australia’s last periodic review, to highlight specific points of divergence between different state and territory anti-vilification laws, which are such as to have a potentially serious impact on the ability of the public to

predict what they can and cannot say. Part IV considers how well Australian legislation matches guidance for implementing UN standards, for this, after all, is ultimately the issue that Australia will face at its next review before the Committee. To that end, Part V considers what might be an appropriate legislative model, in view of the requisite properties of laws, established at international law, by which freedom of expression may permissibly be restricted. It assesses the impact of certain structural aspects of domestic anti-vilification legislation, in particular the 'categorical' approach to prohibition and exclusion. It also touches on the effect on free speech of intervention by competent authorities. Part VI offers specific legislative proposals, before some observations are made by way of conclusion in Part VII.

II HATE SPEECH MECHANISMS UNDER INTERNATIONAL LAW AND AUSTRALIAN ANTI-VILIFICATION LEGISLATION CONTRASTED

The term 'hate speech' is used by the Human Rights Committee and the CERD Committee, in reference to *ICCPR* art 20 and *ICERD* art 4(a) respectively. It has a specific meaning, though these provisions differ from each other. *ICCPR* art 20(2) requires States Parties to prohibit the 'advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence'.¹⁰ *ICERD* art 4(a) mandates the criminal prohibition of 'all dissemination of ideas based on racial superiority or hatred, incitement to racial discrimination, as well as all acts of violence or incitement to such acts against any race or group of persons of another colour or ethnic origin'.¹¹ Both provisions are consistent with *ICCPR* art 19, and must be implemented in such a way that preserves freedom of expression in accordance with the standards set by art 19.¹² Article 19(3) only permits restrictions on freedom of expression where necessary in support of 'the rights or reputations of others ... the protection of national security or of public order ... or of public health or morals'. *ICERD* art 4(a) and *ICCPR* art 20 each describe a category of speech that is so harmful in its effects that it can confidently be taken to warrant restriction in accordance with the terms of art 19(3), without individualised assessment, but in the specific terms defined in art 4(a) and 20 respectively. Hate speech restrictions may be allowed outside those categories, or

¹⁰ On the origin of art 20(2), see Jeroen Temperman, *Religious Hatred and International Law: The Prohibition of Incitement to Violence or Discrimination* (Cambridge University Press, 2015) chs 2, 3.

¹¹ See *ibid* 128.

¹² UN Human Rights Committee, *General Comment No 11 (1983): Prohibition of Propaganda for War and Inciting National, Racial or Religious Hatred (Article 20 of the International Covenant on Civil and Political Rights)*, UN Doc CCPR/C/GC/11 (29 July 1983) [2]. See also Paul Taylor, *A Commentary on the International Covenant on Civil and Political Rights: The UN Human Rights Committee's Monitoring of ICCPR Rights* (Cambridge University Press, 2020) 580; Aroney and Taylor, 'The Rights and Wrongs of 18C' (n 9), Part II.1.

at lower thresholds, but only if they meet the standards of justification set by art 19(3).¹³

One important observation should be made at this point on *ICERD* art 4(a) and *ICCPR* article 20. On their face, these provisions address very narrow grounds ('national, racial or religious hatred'), but as the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (Special Rapporteur) has indicated, the prohibitions of incitement that they embody may also be understood to apply to broader categories than race and religion, matching the prohibited grounds of discrimination now acknowledged under the *ICCPR* (covering race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status, including indigenous origin or identity, disability, migrant or refugee status, sexual orientation, gender identity or intersex status, age and albinism).¹⁴

The civil prohibitions in Australian anti-vilification legislation are not confined to the scope of prohibitions mandated by *ICERD* art 4(a) and *ICCPR* art 20. They do not specifically target 'hate speech' as that term is used by the UN monitoring bodies. 'Vilification' is the more apt description. The legislation was not developed to preserve freedom of expression under art 19, and does not incorporate art 19(3) criteria to justify restrictions. 'Freedom of expression' is not, as such, guaranteed in Australia, as a matter of substantive law. A distinction is therefore made in this article between 'freedom of expression', referring to art 19 protection, and 'free speech', as generally available in Australia. The basis of such 'free speech' as currently exists in Australia subsists in four sources:

- The constitutionally implied freedom of political communication. This could be used to challenge the validity of a law that impermissibly burdens the implied freedom.¹⁵ No claim on this basis has yet succeeded in impugning Australian anti-vilification legislation. The implied freedom applies particular constitutional-law criteria, primarily directed at the exercise of legislative power, and has limited capacity to correct any mis-targeting of legislation that does not transgress that freedom.¹⁶

¹³ On the relationship between arts 19 and 20, see Nazila Ghanea, 'Expression and Hate Speech in the ICCPR: Compatible or Clashing?' (2010) 5(2-3) *Religion and Human Rights* 171; Nazila Ghanea, 'Intersectionality and the Spectrum of Racist Hate Speech: Proposals to the UN Committee on the Elimination of Racial Discrimination' (2013) 35(4) *Human Rights Quarterly* 935, 935-8.

¹⁴ David Kaye, *Report of the Special Rapporteur on the Promotion of the Right to Freedom of Opinion and Expression*, Un Doc A/74/486 (9 October 2019) 6 [9] ('Special Rapporteur Report A/74/486').

¹⁵ See Nicholas Aroney, 'The Constitutional (In)Validity of Religious Vilification Laws: Implications for Their Interpretation' (2006) 34(2) *Federal Law Review* 287.

¹⁶ For example, the Supreme Court of Tasmania (Brett J) rejected a constitutional challenge to s 17 of the *Anti-Discrimination Act 1998* (Tas) in *Durston v Anti-Discrimination Tribunal* (No 2) [2018] TASSC 48.

- The common law notion that ‘everybody is free to do anything, subject only to the provisions of the law’ (such provisions include anti-vilification legislation).¹⁷
- The principle of legality, which requires courts not to impute to the legislature an intention to interfere with fundamental rights and freedoms, including freedom of speech, in the absence of an intention that is clearly manifested by unmistakable and unambiguous language.¹⁸
- The interpretive mandates of the human rights charters of the ACT, Victoria and Queensland. These charters require statutory provisions to be interpreted in a way that is compatible with listed human rights (approximating those in the ICCPR), so far as it is possible to do so consistently with the purpose of such statutory provisions.¹⁹ However, they do not give substance to the ICCPR concept of freedom of expression within the categories of exclusion from the anti-vilification prohibitions.

State and territory anti-vilification statutes remain the primary basis on which vilification claims are made and contested. They strike a particular balance between the rights affected. This is achieved by closely defined categories of prohibited and excluded speech. However, it may be that too much is expected of these laws in adequately differentiating between the harmful speech that is properly its prime target, and other speech that ought to remain untouched.

The important point to note at this stage is that there are material divergences between the mechanisms in Australian law just described and the standards of justification for limitations on freedom of expression required by art 19. The *ICCPR* requires States Parties such as Australia ‘to adopt such laws and other measures as may be necessary to give effect to the rights recognised in the [*ICCPR*]’.²⁰ This includes freedom of expression.

The absence of any substantive *ICCPR* protection for freedom of expression widens the scope for Australian jurisdictions to formulate their own anti-vilification standards. The regulation of harmful speech is a politically sensitive area in which policy responses are devised to meet the social conditions understood to exist in each jurisdiction at a particular point in time. The result is varying standards across jurisdictions. There is also the possibility that competent authorities will interpret and apply the legislation in a way that introduces further variance.

¹⁷ See, eg, *Coleman v Power* (2004) 220 CLR 1, 97–8 [253] per Kirby J: ‘everybody is free to do anything, subject only to the provisions of the law’.

¹⁸ *Coco v The Queen* (1994) 179 CLR 427, 437 (Mason CJ, Brennan, Gaudron and McHugh JJ).

¹⁹ *Charter of Human Rights and Responsibilities Act 2006* (Vic) s 32(1); *Human Rights Act 2019* (Qld) s 48; *Human Rights Act 2004* (ACT) s 30.

²⁰ *ICCPR* (n 6) art 2(2).

We now turn in Part III to the detail of state and territory anti-vilification legislation so that it may be assessed in Part IV against the international standards and guidance from UN sources.

III AUSTRALIAN ANTI-VILIFICATION LEGISLATION

There is significant disparity across Australian jurisdictions in their approaches to vilification. The multiplicity of statutes at Commonwealth, state and territory levels, with readily observable divergences between them, renders Australia's anti-vilification laws among the most complicated in the world. The following is an attempt to simplify matters by identifying key similarities and differences across Australian jurisdictions.²¹

There is a pattern in most states and territories of two-tiered prohibition against vilifying conduct through the establishment of criminal and civil prohibitions and sanctions. The legislation has the following prominent features.

Criminal provisions apply to what is often termed 'serious vilification'. These offences have two key elements: first, incitement of hatred against or towards, serious contempt for, or severe ridicule of, a person or group; and second, that incitement involves threats of physical harm or damage to property (or incitement of others to such harm or damage).²² Within this category, there are variations across jurisdictions:

- as to applicable grounds: race (South Australia, Western Australia); race and religious belief or activity (Victoria); race, religion, sexuality or gender identity (Queensland); race, religion, sexual orientation, gender identity or intersex or HIV/AIDS status (New South Wales); and disability, gender identity, HIV/AIDS status, race, religious conviction, sex characteristics, sexuality (ACT);
- as to aggravating factors required to be proved: inciting or threatening violence need not be proved in Western Australia;
- as to the conduct described: by additional terms such as 'revulsion' (ACT and Victoria); by the brevity of some offence definitions (ACT, New South Wales);²³ by added qualifications, eg to render irrelevant any

²¹ A useful comparison was made between the NSW and Commonwealth models in Katharine Gelber and Luke McNamara, 'Anti-Vilification Laws and Public Racism in Australia: Mapping the Gaps between the Harms Occasioned and the Remedies Provided' (2016) 39(2) *University of New South Wales Law Journal* 488, 497–8. For discussion on when different grounds of prohibited vilification were introduced, and the different thresholds, see Katharine Gelber and Luke McNamara, 'The Effects of Civil Hate Speech Laws: Lessons from Australia' (2015) 49(3) *Law and Society Review* 631, 634–7.

²² *Criminal Code 2002* (ACT) s 750; *Crimes Act 1900* (NSW) s 93Z; *Anti-Discrimination Act 1991* (Qld) s 131A; *Racial and Religious Tolerance Act 2001* (Vic) ss 24–25; *Racial Vilification Act 1996* (SA) s 4; *Criminal Code Act Compilation Act 1913* (WA) ss 77–80H.

²³ *Criminal Code 2002* (ACT) s 750; *Crimes Act 1900* (NSW) s 93Z.

assumptions about an attribute and whether any one was actually incited (New South Wales);²⁴ in terminology used (in Western Australia the key phrases are ‘animosity towards’ (meaning hatred of or serious contempt for), and ‘harassment’ (which includes to threaten, seriously and substantially abuse or severely ridicule)); and in the harm addressed (in Western Australia additional offences are committed by the possession of material for dissemination or display with that purpose);

- as to the penalties that apply: these include fines varying in severity; custodial sentences in some jurisdictions with maxima of six months (Queensland) or three years (New South Wales),²⁵ and in some jurisdictions no custodial sentencing.²⁶

Civil provisions apply to less serious ‘vilification’. These provisions target a particular form of conduct typically (but not invariably) defined as inciting hatred against, serious contempt for, or severe ridicule of, a person or group, without requiring threats of physical harm or damage to property.²⁷ There are much more significant variations across jurisdictions in civil than criminal provisions:

- as to grounds: race and religious belief or activity (Victoria); race, transgender, sexual orientation and HIV/AIDS (New South Wales); race, religion, sexuality or gender identity (Queensland); disability, gender identity, HIV/AIDS status, race, religious conviction, sex characteristics, sexuality (ACT); race, disability, sexual orientation, lawful sexual activity, religious belief or affiliation, religious activity, gender identity or intersex variations of sex characteristics (Tasmania’s s 19);
- as to the conduct described: by additional terms such as ‘revulsion’ (ACT and Victoria); or a different formula (‘offended, humiliated, intimidated, insulted or ridiculed’ in Tasmania’s s 17 ‘prohibition of certain conduct’;²⁸ ‘offend, insult, humiliate or intimidate’ in Northern Territory’s s 20A prohibition of ‘offensive behaviour because of attribute’);²⁹
- as to the areas of activity to which it applies: the Tasmanian ‘prohibition of certain conduct’,³⁰ and the Northern Territory’s prohibition of

²⁴ *Crimes Act 1900* (NSW) s 93Z.

²⁵ See, eg, *Anti-Discrimination Act 1991* (Qld) s 131A(1) (70 penalty units or 6 months imprisonment). Cf *Crimes Act 1900* (NSW) s 93Z(1) (100 penalty units or imprisonment for 3 years, or both).

²⁶ *Criminal Code 2002* (ACT) s 750 (maximum penalty: 50 penalty units, no imprisonment).

²⁷ *Discrimination Act 1991* (ACT) s 67A; *Anti-Discrimination Act 1977* (NSW) ss 20C(1), 38S, 49ZT, 49ZXB; *Racial and Religious Tolerance Act 2001* (Vic) ss 7–8; *Anti-Discrimination Act 1991* (Qld) s 124A; *Anti-Discrimination Act 1998* (Tas) ss 17, 19; *Civil Liability Act 1936* (SA) s 73(1).

²⁸ *Anti-Discrimination Act 1998* (Tas) s 17(1).

²⁹ *Anti-Discrimination Act 1992* (NT) s 20A(1).

³⁰ *Anti-Discrimination Act 1998* (Tas) ss 17(1), 22.

‘offensive behaviour because of attribute’,³¹ apply to the same areas of activity as the general anti-discrimination provisions;³²

- as to the range of eligible exclusions, which generally allow, though in different terms: fair reporting; academic, artistic, scientific or research and other purposes in the public interest (including discussion or debate); and any matter which is subject to a defence of absolute privilege in a proceeding for defamation, typically in parliamentary or court/tribunal proceedings (ACT, New South Wales, Queensland, Tasmania and South Australia), with an additional exception in Victoria for the performance, exhibition or distribution of an artistic work;
- as to the conceptual basis for liability: South Australia relies on the tort of ‘racial victimisation’;³³
- as to the threshold: the Tasmanian ‘prohibition of certain conduct’,³⁴ and the Northern Territory’s prohibition of ‘offensive behaviour because of attribute’,³⁵ apply a standard normally associated with sexual harassment.³⁶

The jurisdictions also vary as to whether their legislation adopts:

- both civil and criminal provisions (ACT, New South Wales, Queensland, Victoria, South Australia);
- civil but not criminal provisions (Tasmania and Northern Territory);
- more than one civil vilification regime (in Tasmania, one section prohibits conduct which offends, humiliates, intimidates, insults or ridicules; the other prohibits ‘inciting hatred’);³⁷
- criminal but not civil provisions (Western Australia).

Given all these differences, the safest approach to avoiding liability everywhere in Australia is to heed the highest standard of prohibition found in any Australian jurisdiction. In claims based on social media posts, it is likely that jurisdiction may be asserted in any state or territory in the face of evidence that the offending posts were accessed there.³⁸

Like the term ‘hate speech’, when used generically (rather than as defined in UN instruments) the term ‘vilification’ lacks precise meaning, although the common ground already traversed is a starting point. ‘Serious vilification’ may be

³¹ *Anti-Discrimination Act 1992* (NT) s 20A(1), 28.

³² *Ibid* ss 17(1), 22.

³³ *Civil Liability Act 1936* (SA) s 73.

³⁴ *Anti-Discrimination Act 1998* (Tas) s 17(1).

³⁵ *Anti-Discrimination Act 1992* (NT) s 20A(1).

³⁶ *Anti-Discrimination Act 1998* (Tas) s 17(1).

³⁷ *Ibid* ss 17, 19.

³⁸ *Clinch v Rep* [2020] ACAT 13, 7–10 [13]–[23].

taken to denote the *criminal* standard, although this description applies to only half of the jurisdictions that have enacted criminal provisions (other descriptions rely on the phraseology ‘publicly threatening or inciting violence’ on specified grounds (New South Wales), racial vilification (South Australia), and ‘racial animosity,’ or ‘racist harassment’ (Western Australia)). Most of the jurisdictions that include *civil* provisions use the term ‘vilification’ in different ways, but alternative descriptions include ‘victimisation’ (South Australia) or ‘inciting hatred’ (Tasmania, s 19), when adopting the same civil formula of ‘inciting hatred, serious contempt or severe ridicule.’ Section 18C of the *Racial Discrimination Act 1975* (Cth) (‘RDA’) uses the term ‘offensive behaviour’.³⁹

Tasmania’s s 17(1) is particularly difficult to characterise, as is the Northern Territory’s s 20A. Section 17(1) prohibits conduct that ‘offends, humiliates, intimidates, insults or ridicules’ another on the basis of a qualifying attribute, in circumstances in which a reasonable person, having regard to all the circumstances, would have anticipated that the other person would be offended, humiliated, intimidated, insulted or ridiculed. The Northern Territory’s test is whether the act ‘is reasonably likely, in all the circumstances, to offend, insult, humiliate or intimidate another person or a group’. Both provisions differ substantially from conventional state and territory anti-vilification legislation for including the words, ‘offends’, ‘insults’ and ‘ridicules’. The terms ‘hatred’, ‘serious contempt’ and ‘severe ridicule’ are used in most other civil and criminal vilification prohibitions and also in s 19 of the Tasmania law. Tasmania’s 17(1), and the Northern Territory’s s 20A, apply in a more limited way than typical vilification prohibitions, in specified areas of activity, within the same confines as the legislation’s discrimination (and harassment) provisions. For example, all that is required under the Tasmanian legislation is that the conduct occurs ‘by or against a person engaged in, or undertaking any, activity in connection with’ any specified areas of activity, such as employment or provision of facilities, goods or services.⁴⁰ Moreover, there is no requirement that the s 17(1) conduct must occur in public (or to similar effect), unlike all other Australian anti-vilification legislation.

Tasmania’s 17(1), and the Northern Territory’s s 20A, differ from conventional harassment legislation in a number of respects. All states and territories have sexual harassment legislation in similar form and with specific, elaborated meaning (including Tasmania, in s 17(2)–(3), and the Northern Territory in s 22). The general formula for sexual harassment throughout Australia has been the model established by the *Sex Discrimination Act 1984* (Cth) s 28A, of ‘offended, humiliated or intimidated’, which omits ‘insult and ridicule’ found in Tasmania’s s 17(1) and ‘insult’ found in the Northern Territory’s s 20A.⁴¹

³⁹ See Aroney and Taylor, ‘The Rights and Wrongs of 18C’ (n 9).

⁴⁰ *Anti-Discrimination Act 1998* (Tas) s 22(1); *Anti-Discrimination Act 1002* (NT) s 28.

⁴¹ For further discussion, see Gus Bernardi, ‘From Conflict to Convergence: The Evolution of Tasmanian Anti-Discrimination Law’ (2001) 7(1) *Australian Journal of Human Rights* 134.

If treated as an anti-harassment measure, s 17(1) is unusual for legislating beyond the realm of sexual harassment, though some jurisdictions do so, to a limited extent.⁴² Tasmania's s 17(1) thus has associations with harassment legislation but it is not confined to that realm. It also possesses characteristics linked with vilification.⁴³ The way it has been applied by the competent authority, as discussed in Part V below, means that it must be characterised as an anti-vilification measure (especially since ss 17 and 19 are subject to the same vilification exceptions in s 55). The Northern Territory's definition of discrimination in s 20(1) includes harassment on the basis of an attribute. As harassment provisions they are therefore broad. As vilification provisions they are the very broadest of their kind in Australia.⁴⁴

IV ASSESSMENT AGAINST INTERNATIONAL STANDARDS

This Part describes the guidance provided by UN sources on implementing hate speech provisions required by *ICERD* and the *ICCPR*, ranging from the most serious categories (for which *ICERD* art 4(a) and *ICCPR* art 20(2) mandate prohibition), beyond which the general principles of art 19(3) require specific justification. Australian anti-vilification legislation is assessed against that guidance.

A *ICERD Article 4(a) Criminal Provisions*

According to the CERD Committee's General Recommendation 35, *Combating Racist Hate Speech*, criminal prohibition should be reserved only for serious cases of racist expression, while less serious conduct should be addressed by other means, taking into account such things as the nature and extent of the impact on targeted persons and groups.⁴⁵ The CERD Committee cited the Human Rights Committee's *General Comment No 34 on Freedom of Expression*, to stress that criminal sanctions should always be governed by principles of legality, proportionality and necessity. The CERD Committee also recommended (in reference to *ICERD* art 4(a)) that States Parties declare and effectively sanction as offences punishable by law:

⁴² *Equal Opportunity Act 1984* (WA) ss 49A–C; *Anti-Discrimination Act 1992* (NT), ss 19–20. Sexual harassment is the subject of s 22.

⁴³ Some anti-vilification measures describe themselves in terms of harassment (Western Australia's criminal provisions).

⁴⁴ At the time of writing, the Queensland Government is considering introduction of a provision that would make a criminal offence to publicly display a prohibited symbol 'in a way that might reasonably be expected to cause a member of the public to *feel* menaced, harassed or *offended*': Criminal Code (Serious Vilification and Hate Crimes) Amendment Bill 2023 proposing insertion of a new s 52D into the *Criminal Code Act 1899* (Qld).

⁴⁵ See Temperman (n 10) 128.

- (b) Incitement to hatred, contempt or discrimination against members of a group on grounds of their race, colour, descent, or national or ethnic origin;
- (c) Threats or incitement to violence against persons or groups on [those grounds];
- (d) Expression of insults, ridicule or slander of persons or groups or justification of hatred, contempt or discrimination on [those grounds], when it clearly amounts to incitement to hatred or discrimination.⁴⁶

In Australia, the key elements of ‘serious vilification’ broadly correspond with the *ICERD* Committee’s description, where they concern inciting hatred against or towards, serious contempt for, or severe ridicule of, a person or group, involving threats of physical harm or damage to property (or incitement of others to such harm or damage). There are no defences equivalent to the exclusions in Australian civil prohibitions (i.e., academic, artistic, scientific or research purposes, discussion or debate, and fair reporting), but so long as the offences are defined so they attach only to the most serious hate speech, involving threats of physical harm or damage to property, this poses much less of a problem than the civil prohibitions. It is, however, important to note the gloss put on the CERD Committee’s recommendations by the Special Rapporteur, David Kaye, to prevent even *ICERD*’s criminal provisions from being interpreted as supporting greater restrictions on freedom of expression than is allowed under the *ICCPR*. Noting the fundamental nature of the freedom of expression, he pointed out that restrictions on the freedom ‘must be exceptional, subject to narrow conditions and strict oversight’.⁴⁷ The standards of art 19(3) must therefore be strictly preserved, he said, when implementing *ICERD*’s criminal requirements.⁴⁸ Noting that the term ‘ridicule’ is a very broad term, he emphasised that it must only be prohibited where it ‘clearly amounts to incitement to hatred or discrimination’. Indeed, ridicule is a type of speech ‘generally precluded from restriction under international human rights law, which protects the rights to offend and mock’. Thus, prohibitions of such categories of expression must be limited to the most serious cases of hate speech. For this reason, ‘the ties to incitement and to the framework established under article 19(3) of the Covenant help to constrain such a prohibition to the most serious category’.⁴⁹ Article 19(3) delineates an area of exclusion that should apply to all Australian anti-vilification legislation, criminal and civil, if it is to comply with the *ICCPR*.

Similar findings were made in a series of expert workshops on the inter-relationship between freedom of expression and hate speech known as the *Rabat Plan of Action*, which was firmly endorsed by the UN High Commissioner for

⁴⁶ Committee on the Elimination of Racial Discrimination, *General Recommendation No 35 (2013), Combating Racist Hate Speech*, UN Doc CERD/C/GC/35 (26 September 2-13) [12]–[13] (‘GR 35’).

⁴⁷ Special Rapporteur Report A/74/486 (n 14) 5 [5]–[6].

⁴⁸ *Ibid* 8 [16].

⁴⁹ *Ibid* [16]–[17].

Human Rights when it was launched in 2013.⁵⁰ Both the *Rabat Plan of Action* and the CERD Committee in its General Recommendation identified certain factors denoting the severity necessary to criminalise ‘incitement’.⁵¹ The CERD Committee made the following observations on this element:

Incitement characteristically seeks to influence others to engage in certain forms of conduct, including the commission of crime, through advocacy or threats. Incitement may be express or implied, through actions such as displays of racist symbols or distribution of materials as well as words. The notion of incitement as an inchoate crime does not require that the incitement has been acted upon, but in regulating the forms of incitement referred to in article 4, States parties should take into account, as important elements in the incitement offences, in addition to [the content and form of speech; the economic, social and political climate; the position or status of the speaker; the reach of the speech; the objectives of the speech], the intention of the speaker, and the imminent risk or likelihood that the conduct desired or intended by the speaker will result from the speech in question.⁵²

Australian *criminal* anti-vilification provisions may be said to match broadly these requirements of *ICERD*, though implementation in terms of *ICERD* art 4(a) offers obvious advantages by way of international law compliance and predictability through uniformity, and a means of correcting divergence from that standard where it exists.

B ICCPR Article 20(2)

The *Rabat Plan of Action* is especially important to our present discussion because it attempted to clarify some difficult definitions found in *ICCPR* art 20 and made specific recommendations for domestic implementation. It was influenced by the *Camden Principles on Freedom of Expression and Equality* (‘*Camden Principles*’) developed by the non-governmental organisation ‘Article 19’.⁵³ The *Rabat Plan of Action* concluded that, for the purposes of art 20:

⁵⁰ Human Rights Council, *Annual Report of the United Nations High Commissioner for Human Rights, Addendum Report of the United Nations High Commissioner for Human Rights on the Expert Workshops on the Prohibition of Incitement to National, Racial or Religious Hatred*, UN Doc A/HRC/22/17/Add.4 (11 January 2013) (‘A/HRC/22/17/Add.4’).

⁵¹ A/HRC/22/17/Add.4, n 45, Appendix, *Rabat Plan of Action on the Prohibition of Advocacy of National, Racial or Religious Hatred that Constitutes Incitement to Discrimination, Hostility or Violence* 11 [29] (‘*Rabat Plan of Action*’); GR 35 (n 46) 4–5 [12]–[15].

⁵² GR 35 (n 46) 5 [16]. This draws on both the Human Rights Committee’s *General Comment No 34, Article 19: Freedoms of Opinion and Expression*, UN Doc CCPR/C/GC/34 (12 September 2011) (‘GC 34’) and the *Rabat Plan of Action* (n 51).

⁵³ Article 19, *The Camden Principles on Freedom of Expression and Equality* (Report, April 2009) <<https://www.article19.org/data/files/pdfs/standards/the-camden-principles-on-freedom-of-expression-and-equality.pdf>> (‘*Camden Principles*’).

- ‘Hatred’ and ‘hostility’ refer to ‘intense and irrational emotions of opprobrium, enmity and detestation towards the target group.’
- ‘Advocacy’ is to be understood as requiring ‘an intention to promote hatred publicly towards the target group.’
- ‘Incitement’ refers to statements which create an imminent risk of discrimination, hostility or violence against persons belonging to target groups.⁵⁴
- ‘Incitement to hatred’ refers to the ‘most severe and deeply felt form of opprobrium,’ where relevant elements informing the severity of the hatred may include the ‘cruelty or intent of the statement or harm advocated, the frequency, quantity and extent of the communication’.
- There must be ‘intent’ to incite; negligence and recklessness are not sufficient.⁵⁵

It is arguable that this last requirement was not adhered to in the interpretation of the term ‘incitement’ in s 49ZT of the *Anti-Discrimination Act 1977* (NSW) adopted by Bathurst CJ in *Sunol v Collier (No 2)*.⁵⁶ His Honour considered that it is not necessary for a person to be incited by the words or publication, nor is it necessary to establish an intention to incite.⁵⁷ In *Cottrell v Ross*,⁵⁸ it was similarly held that, under s 8 of the *Racial and Religious Tolerance Act 2001* (Vic), ‘the intention of the inciter to incite is irrelevant to the question of liability’.⁵⁹

The *Rabat Plan of Action* supported the *Camden Principles’* recommendation that ICCPR States Parties ‘should make it clear, either explicitly or through authoritative interpretation’, that these definitions and requirements apply to any law implementing art 20 so that an appropriately high threshold for unlawful conduct is maintained.⁶⁰

Both CERD and the *Rabat Plan for Action* treat as important the social and political setting, and the nature and extent of the impact of the conduct, which the *Rabat Plan for Action* expressed in terms of ‘imminence’, meaning ‘the courts will have to determine that there was a reasonable probability that the speech

⁵⁴ *Rabat Plan of Action* (n 51) 9–10 [21], n 5.

⁵⁵ *Ibid* 11 [29].

⁵⁶ (2012) 260 FLR 414 (‘*Sunol v Collier*’).

⁵⁷ *Ibid* 424 [41(a)] (Bathurst CJ). See also *Margan v Manias* [2015] NSWCA 388 [11]–[15]. As Temperman observes, the element of ‘incitement’ in art 20(2) requires that the speaker *intends to incite* a third party to engage in discrimination, hostility or violence against a target group: Temperman (n 10) 180–1, 209–14, 237–8.

⁵⁸ [2019] VCC 2142.

⁵⁹ *Ibid* 8 [42]. See also *Catch The Fire Ministries Inc v Islamic Council of Victoria Inc* [2006] VSCA 284, 9–10 [23]–[24].

⁶⁰ *Rabat Plan of Action* (n 51) [21], n 5, citing Camden Principle 12. See also Manfred Nowak, *U.N. Covenant on Civil and Political Rights: CCPR Commentary* (Engel, 2nd ed, 2005) 475; Temperman, (n 10) 164; Aroney and Taylor ‘The Rights and Wrongs of 18C’ (n 9).

would succeed in inciting actual action against the target group, recognizing that such causation should be rather direct.’ Other factors include the status of the speaker, the extent to which the speech was provocative and direct, and the size of the audience.⁶¹

Although the *Rabat Plan for Action* is directed to legislation specifically implementing art 20(2), it provides authoritative guidance concerning the kinds of measures addressing hate speech that will conform to art 19(3) requirements. When measured against art 20(2), it is apparent that the Australian state and territory civil prohibitions do not include the element of ‘incitement to discrimination, hostility or violence’. The Victorian civil provisions, for example, require that the conduct ‘incites hatred against, serious contempt for, or revulsion or severe ridicule’ of another person or class of persons.⁶² They were considered by the Victorian Court of Appeal in *Catch The Fire Ministries Inc v Islamic Council of Victoria Inc*.⁶³ Whereas the *Rabat Plan for Action* understands hatred and hostility to refer to ‘intense and irrational emotions of opprobrium, enmity and detestation’, there was no consideration in the appeal judgment of the seriousness of the hatred required by the Victorian law except to draw attention to the statutory language itself and Neave JA’s observation that the law is concerned with incitement of ‘extreme responses’.⁶⁴ The *Rabat Plan for Action* says that incitement refers to statements that create an *imminent risk* of discrimination, hostility or violence; Neave JA appeared to consider it sufficient that the conduct be merely capable of causing hatred.⁶⁵

C ICCPR Article 19(3)

In his 2019 report on the regulation of online hate speech, the Special Rapporteur expressed some caution about adopting even the established terminology of ICCPR art 20(2) and ICERD art 4(a) in anti-vilification legislation, because of its uncertainty. He considered the language of these convention provisions to be ambiguous, dependent on difficult-to-define language of emotion (hatred, hostility) and highly context-specific prohibition (advocacy of incitement). They lack precision and require interpretation. His point was to emphasise that, quite apart from the requirements of arts 4(a) and 20(2), whenever a State limits expression, even when implementing art 20(2), it must always ‘justify the prohibitions and their provisions in strict conformity with article 19’.⁶⁶

⁶¹ *Rabat Plan of Action* (n 51) 9–10 [29].

⁶² *Racial and Religious Tolerance Act 2001* (Vic) ss 7–8.

⁶³ *Catch The Fire Ministries* (n 59).

⁶⁴ *Ibid* 54 [174].

⁶⁵ *Ibid* 50 [154], 58 [194], although elsewhere her Honour seemed to indicate that the impugned conduct must be ‘likely to incite hatred’: at 57–8 [190].

⁶⁶ Special Rapporteur Report A/74/486 (n 14) 7–8 [12]–[13], referring to GC 34 (n 52) 12–13 [50]–[52].

Australian anti-vilification prohibitions certainly recruit the phraseology of 'incitement' and 'hatred', with their attendant uncertainty, but with nothing as effective as art 19(3) to save the prohibitions from excess.

The recent *United Nations Strategy and Plan of Action on Hate Speech*, quoted at the outset, reasoned similarly. It opened by noting that the characterisation of what is 'hateful' is controversial and disputed. It proposed as a working definition for hate speech 'any kind of communication ... that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are, in other words, based on their religion, ethnicity, nationality, race, colour, descent, gender or other identity factor'. But it went on to explain that international law does not prohibit hate speech as such, only *incitement* to discrimination, hostility and violence, an especially dangerous form of speech because it explicitly and deliberately aims to trigger discrimination, hostility and violence, which may also lead to or include terrorism or atrocity crimes.⁶⁷

Commenting on this, the Special Rapporteur accepted that the description given to 'hate speech' in the *Strategy and Plan of Action* is appropriate as a basis for political and social action to counter discrimination and hatred. However, given its vagueness it would be problematic as the basis for legislative prohibitions, on legality grounds (discussed further below). There is a crucial distinction between conduct deserving a political, social or educative response, and conduct warranting legislative curtailment. To the Special Rapporteur, it remained essential that the State demonstrate the necessity and proportionality of legislative action restricting freedom of expression. The harsher the penalty, the greater the need to demonstrate the necessity for the measure in exacting terms.⁶⁸ Of particular relevance to the civil prohibitions in anti-vilification legislation, the Special Rapporteur also stressed that:

expression that may be offensive or characterized by prejudice and that may raise serious concerns of intolerance may often not meet a threshold of severity to merit any kind of restriction. There is a range of expression of hatred, ugly as it is, that does not involve incitement or direct threat, such as declarations of prejudice against protected groups. Such sentiments would not be subject to prohibition under the [ICCPR] or [ICERD], and other restrictions or adverse actions would require an analysis of the conditions provided under article 19 (3) of the Covenant.⁶⁹

The main lesson to be applied to Australian anti-vilification provisions is that concepts that defy precise definition may be tolerated in prohibitions, *provided* their ill-effects can be rescued by the operation of principles that give substance to art 19(3).

⁶⁷ *Strategy and Plan of Action* (n 1) 2.

⁶⁸ Special Rapporteur Report A/74/486 (n 14) 9–10 [20].

⁶⁹ *Ibid* [24].

V LEGISLATIVE MODELLING

In this Part, we identify the properties that laws must possess if they are to restrict freedom of expression permissibly under art 19(3). We raise concerns about the administration of the law by equal-opportunity, anti-discrimination and human-rights commissions in Australia (remembering that the requirements of art 19(3) attach to restrictions imposed by courts, public authorities and others when interpreting and applying the law in practice), and we discuss the impact of particular legislative models, particularly the shortcomings of the ‘categorical’ approach taken in Australia to legislative prohibition and exclusion.

A *Requisite Properties of the Law*

The expectations set by art 19(3) may be summarised in this way. All restrictions on freedom of expression are to be ‘provided by law’ and must be ‘necessary’ for a specified purpose, namely to protect ‘the rights and reputations of others’, ‘national security or ... public order (ordre public), or ... public health or morals’. Any such restriction must be directly related to the specific purpose on which it is predicated, and it must not be overly broad. The restriction must be appropriate to achieve its protective function, it must be the least intrusive means of achieving it, and it must be proportionate to the interest to be protected. This principle of proportionality has to be respected not only in the law that frames the restrictions but also by the administrative and judicial authorities that apply the law.⁷⁰ The law must provide sufficient guidance to those charged with its execution to enable them to ascertain what sorts of expression are properly restricted and what sorts are not.⁷¹ The law must not confer unfettered discretion for the restriction of freedom of expression on those charged with its execution.⁷²

When reporting in 2013 on the *Rabat Plan of Action*, the UN High Commissioner for Human Rights acknowledged that ‘[p]roperly balancing freedom of expression and the prohibition of incitement to hatred is no simple task,’ and he continued, ‘[l]et me state clearly that any limitations to this fundamental freedom must remain within strictly defined parameters flowing from the international human rights instruments, in particular the [ICCPR] and [ICERD]. Article 19, paragraph 3, of the [ICCPR] lays down a clear test by which the legitimacy of such restrictions may be assessed’.⁷³

The *Rabat Plan for Action* itself observed that the broader the definition of incitement to hatred in domestic legislation, the more it opens the door for arbitrary application of the law. It noted an increasing trend towards vagueness

⁷⁰ GC 34 (n 52) 6 [22], 8 [34].

⁷¹ *Ibid* 6–7 [25].

⁷² *Ibid*.

⁷³ A/HRC/22/17/Add.4 (n 50) 4 [9].

in the terminology used in hate speech legislation, and the creation of new categories of restrictions not supportable by reference to art 19(3), or art 20. The limitation of speech effected by such laws must remain an exception to the general principle of freedom of expression.

The requirements of art 19(3) are of general application. They are that:

restrictions are clearly and narrowly defined and respond to a pressing social need; are the least intrusive measure available; are not overly broad, so that they do not restrict speech in a wide or untargeted way; and are proportionate so that the benefit to the protected interest outweighs the harm to freedom of expression, including with respect to the sanctions they authorize.⁷⁴

The greatest risk of Australia's civil prohibitions is that they are capable of restricting speech 'in a wide or untargeted way'.⁷⁵ In part, their breadth is a function of definitional uncertainties that cannot be avoided. At the same time, the Australian exclusions are narrow, admitting only fair reporting, academic, artistic, scientific or research, and other purposes in the public interest, including discussion or debate. These descriptions represent particularised forms of expression in contrast to the comprehensive protection of freedom of expression that should be secured by Australian law. As noted, to the Special Rapporteur 'the freedom of expression defined in article 19(2) involves expansive rights embodied by active verbs (seek, receive, impart) and the broadest possible scope (ideas of all kinds, regardless of frontiers, through any media)'.⁷⁶

B *Legal Certainty*

What matters for ICCPR compliance is whether Australia's civil prohibitions are appropriately targeted to prevent the harms of hate speech and give sufficient scope to freedom of expression. State and territory civil prohibitions appear to have been effective in addressing obvious harms in the many extreme instances of status-based vilification, which fully demand enforcement by state and territory authorities, as attested in the extensive tribunal case law. There are many illustrations in decided cases of supremely fatuous, hate actuated speech, without any justification, by those intent on inflicting harm on targeted individuals on the basis of particular attributes. Those who engage in speech of this kind do not have to be well-versed in the law to know that their speech will and should get them into trouble. However, while the professed aims and aspirations of anti-vilification legislation are laudable, it is questionable whether the full scope of the civil prohibitions is warranted having regard to art 19. For example, the Tasmanian legislation rightly 'seeks to prevent and redress conduct

⁷⁴ *Rabat Plan of Action* (n 51) [15]–[18].

⁷⁵ *Ibid.*

⁷⁶ Special Rapporteur Report A/74/486 (n 14) 7 [12].

which is seen as unjust, divisive and anathema to modern society.⁷⁷ However, s 17 of the *Anti-Discrimination Act 1998* (Tas) is of special concern, particularly in the light of the recent action by Equal Opportunity Tasmania against Tasmanian senator Claire Chandler, following her comments in an opinion piece on women in sport and the use of change-rooms. These appeared in July 2020 in *The Mercury*:

Compounding this inequity [of free speech being reserved for those such as JK Rowling who have the resources or the platform to defend themselves] is the deference that is increasingly demanded to views which ordinary people, whether on the left or right, find absurd and completely lacking in evidence. You don't have to be a bigot to recognise the differences between the male and female sexes and understand why women's sports, single sex changerooms and toilets are important. The overwhelming majority of the world's population grew up understanding these concepts.⁷⁸

The next day a complaint was filed with Equal Opportunity Australia by a constituent who then contacted Senator Chandler by email and asked whether she knew the difference between sex and gender. Senator Chandler replied, 'I do understand the difference – that's why I've made the point in my article that women's toilets and women's change rooms are designed for people of the female sex (women) and should remain that way.'⁷⁹ In deciding that possible breaches were disclosed, the assessment decision of Equal Opportunity Australia referred to the email correspondence.⁸⁰ The complaint was later dropped.⁸¹ Professor Anne Twomey commented that exchanges like this, about public policy, between a senator and a constituent 'go to the core' of the Constitution's implied freedom of political communication.⁸² Senator Chandler considered that she was merely advocating the sex-based rights of women. Reflecting on what had occurred, she later commented that 'it is deeply concerning in a democracy, instead of using free speech to respond or perhaps even campaign against me in an election, some people are instead seeking to use the law to silence me and every Tasmanian who shares my concerns'.⁸³ If Senator Chandler's remarks are indeed caught by the Tasmanian legislation it has two major implications.

⁷⁷ *Durston v Anti-Discrimination Tribunal (No 2)* [2018] TASSC 48, [53].

⁷⁸ Claire Chandler, 'Opinion Piece: Big Names Spark Turning Point against Free Speech Attacks', *The Mercury* (online, 20 July 2020) <<https://www.themercury.com.au>>.

⁷⁹ Claire Bickers, 'Senator in Hot Water for "Humiliating" Transgender Changeroom Remarks', *The Mercury* (online, 5 September 2020) <<https://www.themercury.com.au>>.

⁸⁰ Bernard Lane, 'Free Speech "in Play" over Women's Sport', *The Australian* (online, 11 September 2020) <<https://www.theaustralian.com.au/nation/politics/constitutional-concern-on-complaint/news-story/d6978ee1042354736f0bd6fe7684f6fd>>. As that article notes, to the Commissioner, the implication of the email was that it showed that Sen Chandler 'considers people who are born male and then seek to live as a female, should not have access to female toilets, facilities or sport.'

⁸¹ A conciliation process was formally commenced but was discontinued after Sen Chandler refused to sign a confidentiality agreement. Sen Chandler requested the support of a legal advisor at the meeting, but this was refused.

⁸² Cited by Lane (n 80).

⁸³ Emily Jarvie, 'Discrimination Complaint Filed against Tasmanian Liberal Senator Claire Chandler,' *The Advocate* (Online, 4 September 2020).

First, s 17(1) can only apply because s 22(1) is capable of being construed very broadly. Section 22(1) is meant to confine s 17(1) to prohibited conduct ‘by or against a person engaged in, or undertaking any, activity in connection with’ specific areas of activity.⁸⁴

Secondly, if the complaint was accepted at a time when the email had not been made public by Senator Chandler, this is an example of an anti-vilification prohibition applying to something said in private. Section 17(1) would then be apparently unique in Australian anti-vilification legislation in not requiring the impugned conduct to occur in public (or a non-private setting). As Professor Gillian Triggs, when President of the Australian Human Rights Council, once emphasised, Australian hate speech laws like s 18C of the *Racial Discrimination Act 1975* (Cth) only apply to speech uttered ‘in the public arena’. The discrepancy between her pre-delivered and recorded speech caused some controversy of relevance here,⁸⁵ but the point is this: unless clarified, s 17(1) as applied by Equal Opportunity Tasmania to Senator Chandler, may apply to speech uttered in private contexts.

Putting that detail to one side, the chilling effect of such an uncertain law is a function of both its terms and the manner in which it is administered. Equal Opportunity Tasmania admitted the complaint as if the law it administers applied in those circumstances or, at least, that its application could not be ruled out. The ‘public purpose’ exception in s 55 of the Tasmanian legislation must have been construed extremely narrowly by Equal Opportunity Tasmania for Senator Chandler’s comments not to benefit from it, addressing a question of public policy by an elected member of Parliament. All jurisdictions have public purpose, public interest and similar exclusions.⁸⁶ The point is that hate speech is an evil that must be addressed, but vagueness around the meaning and operation of laws targeting hate speech is likely to encourage self-censorship in areas properly to be regarded as falling within the general domain of freedom of expression. This is especially the case for individuals who do not have the capacity to defend themselves publicly in the way that Senator Chandler did, under protection of parliamentary privilege. However, even Senator Chandler was placed in the invidious position of being officially warned that ‘legal action can be taken against any person who uses insulting language towards any person exercising any power under the Anti-Discrimination Act’.⁸⁷ This inevitably deters public debate about important

⁸⁴ The specific areas are: employment, education and training, provision of facilities, goods and services, accommodation, membership and activities of clubs, administration of any law of the State or any State program, awards, enterprise agreements or industrial agreements.

⁸⁵ Professor Triggs’ pre-delivered speech continued, ‘[o]f course, you can say what you like around the kitchen table.’ What she is reported as having said is different: ‘[s]adly, you can say what you like around the kitchen table at home.’ See Tim Blair, ‘People Who Don’t Do Their Own Research,’ *The Daily Telegraph* (Online, 12 April 2017).

⁸⁶ See Part III above, which identifies the relevant provisions containing anti-vilification prohibitions and exclusions.

⁸⁷ Commonwealth, *Parliamentary Debates*, Senate, 6 October 2020, 5145 (Senator Chandler).

issues, including, ironically, about the appropriateness of an official warning by a public authority being directed against *her*. Even public discussion by *others* about a complaint controversially taken up by the authorities may be deterred, by the fact that it is necessary to recall the original remarks in the process, inviting possible further action.⁸⁸

The Tasmanian legislation highlights the capacity for uncertain anti-vilification legislation to be used as a tool for attacking ideological opponents.⁸⁹ In this context it should be recalled that the freedom of opinion within ICCPR art 19 is violated by ‘harassment, intimidation or stigmatization of a person’, by reason of the opinions they may hold. Any form of effort to coerce the holding or not holding of any opinion is prohibited.⁹⁰

C *Impact of Particular Legislative Models*

As already noted, all of the civil anti-vilification provisions in Australia demarcate the prohibitions so that they do not apply to specific categories of conduct, such as fair reporting and statements made for academic, artistic, scientific or research purposes. The prohibitions are carefully worded to avoid the epithets ‘exclusion’ and ‘exemption’⁹¹ (save for those in Victoria and the Commonwealth).⁹² Referring to the New South Wales legislation in *Sunol v Collier (No 2)*, Allsop P explained that this

reflects an attempt by Parliament to weigh the policies of preventing vilification and permitting appropriate avenues of free speech. Subsection (1), which contains the prohibition, and subsection (2), which stipulates that ‘nothing in this section renders unlawful’ certain conduct, ‘should be read together as a coherent provision that makes certain public acts unlawful. Subsection (2) is not a defence; it is a provision which

⁸⁸ For criticisms of the institutional handling of complaints at federal level, see the report of the Parliamentary Joint Committee on Human Rights, *Freedom of Speech in Australia: Inquiry into the Operation of Part IIA of the Racial Discrimination Act 1975 (Cth) and Related Procedures under the Australian Human Rights Commission Act 1986 (Cth)*, 28 February 2017.

⁸⁹ Most Australian jurisdictions provide for political opinion in some form as a protected attribute in their anti-discrimination provisions, but there are exceptions, notably New South Wales and South Australia.

⁹⁰ GC 34 (n 52) 2–3 [9]–[10].

⁹¹ *Discrimination Act 1991* (ACT) s 67A(2) (‘However, it is not unlawful to’); *Anti-Discrimination Act 1977* (NSW) ss 20C(1), 38S, 49ZT, 49ZXB (‘Nothing in this section renders unlawful’); *Anti-Discrimination Act 1991* (Qld) s 124A (‘Subsection (1) does not make unlawful’); *Anti-Discrimination Act 1998* (Tas) s 55 (‘The provisions of section 17(1) and section 19 do not apply if the person’s conduct is –’); *Civil Liability Act 1936* (SA) s 73(1) (‘act of racial victimisation means ... but does not include...’).

⁹² *Racial and Religious Tolerance Act 2001* (Vic) s 11 (‘Exceptions—public conduct. (1) A person does not contravene section 7 or 8 if the person establishes that the person’s conduct was engaged in reasonably and in good faith ...’); *Racial Discrimination Act 1975* (Cth) s 18D (‘18D Exemptions. Section 18C does not render unlawful anything ...’).

assists in the defining of what is unlawful. It attempts to ensure that certain conduct is not rendered unlawful by the operation of subsection (1).⁹³

This legislative design (a category of prohibition, with stated exclusions) entails certain legal or practical consequences.

First, ordinarily the terms that describe the exclusions are not to be construed narrowly. An approach to construction is to be adopted that recognises ‘the high value that the common law (and indeed the legislature) places on freedom of expression’.⁹⁴

Secondly, the burden of proof rests with the party claiming that their conduct falls within the stated exclusions. This is expressly stated in some legislation.⁹⁵ The rationale is that the party relying on an exclusion is better placed, and has the relevant interest, to explain and to provide evidence of their conduct. In reality, in many cases the assertions in anti-vilification claims give rise, in practice, to a strategic or evidentiary burden on the respondent to rebut the inferences raised. If this is not done, then adverse inferences may be drawn by the tribunal against the respondent. In other words, a shift in burden is, to some extent, unavoidable. However, respondents may pay an unexpected cost even though the exclusions (according to Allsop P) do not operate by way of a defence. The appellant in *Passas v Comensoli* had not specifically pleaded the exclusion and could not raise it for the first time on appeal, because of the disadvantage this would incur for the respondent who would have had the opportunity to answer the exclusion if it had been raised earlier.⁹⁶ In this sense, there is a practical separation between the vilification claim and the public interest exclusions. Until recent changes to the *Australian Human Rights Commission Act 1986* (Cth),⁹⁷ the application of exemptions in s 18D was not always considered when determining whether a complaint amounted to unlawful discrimination.⁹⁸ This may also have been a factor in the handling of the complaint against Senator Chandler by Equal Opportunity Tasmania.

A more fundamental question, however, which arises as a matter of international law, is whether the legislative models adopted by the states and territories offend the principle expressed in General Comment 34 that ‘the

⁹³ *Sunol v Collier* (n 56) 427 [60], in reference to s 49ZT.

⁹⁴ *Ibid* 427 [59].

⁹⁵ See, eg, *Racial and Religious Tolerance Act 2001* (Vic) s 11.

⁹⁶ *Passas v Comensoli* [2019] NSWCATAP 298 (18 December 2019) [52]–[57].

⁹⁷ Section 46PH(1) of the *Australian Human Rights Commission Act 1986* (Cth), concerned with termination of complaints, was added in 2017 to clarify that ‘consideration by the President of the question of whether an act, omission or practice is not unlawful discrimination will involve consideration of whether an exemption applies’.

⁹⁸ Paul Karp, ‘Bill Leak Could Have Ended 18C Complaint Earlier, Says Gillian Triggs: Human Rights Commission President Says Cartoonist Did Not Take Chance to Assert Cartoon Drawn in Good Faith,’ *The Guardian* (online, 28 February 2017). (‘It may very well be fair comment, it may very well be in good faith, it may well be part of an artistic exercise, it may well be accurate,’ Gillian Triggs is attributed as saying. ‘All of those things, however ... have to be put by the respondent themselves.’)

relation between right and restriction and between norm and exception must not be reversed'.⁹⁹ The most obvious example of a breach of that principle would be where a State so restricts a right that it impairs the essence of the right. (ICCPR art 5(1) expresses it in the most severe terms, the 'destruction' of the right, but art 5(1) also refers to less extreme action in the limitation of rights to a greater extent than the ICCPR allows.) To be sure, anti-vilification provisions are not destructive in this extreme sense, and free speech remains as available as ever beyond the scope of the prohibitions. This was illustrated in *Cottrell v Ross*, in a finding that only those communications that vilified were burdened.¹⁰⁰

Nevertheless, this does not answer a concern recently mentioned by the Special Rapporteur when he issued the reminder that, in view of the 'exceptional nature' of limitations allowed under art 19(3), 'the burden falls on the authority restricting speech to justify the restriction, not on the speakers to demonstrate that they have the right to such speech'.¹⁰¹ This requirement is difficult to square with the fact that the protection of free speech occurs by dint only of the legislative carveout, even if it avoids terminology such as 'exclusion' or 'exemption'. This draws attention to what is a radical departure of Australian domestic law from the 'rule or exception' principle, namely that the applicable *rule* — the protection of the fundamental human right in question — does not exist anywhere in substantive Australian law. The freedoms of opinion and expression are not substantively guaranteed in accordance with the requirements of art 19. All applicable convention obligations on hate speech to which Australia is bound require full protection for freedom of expression under art 19. Such protection does not exist in Australia. In its place is the free speech presupposed by the common law, a principle that, as discussed in Part II, offers only meagre protection that fails to equate with freedom of expression. Even that inadequate free speech protection is then circumscribed by statutory anti-vilification prohibitions and saved partially by exceptions. No justification is required beyond a matching of conduct with the prohibitions and exclusions. This is how the balance is fixed in Australia between fundamental rights. It requires little regard for whether the resulting restriction is justified in the circumstances.

It is a cardinal principle of international human rights law that no one human right is to be favoured over another. The *Vienna Declaration and Programme of Action* expressed it in this way:

All human rights are universal, indivisible and interdependent and interrelated. The international community must treat human rights globally in a fair and equal manner, on the same footing, and with the same emphasis. While the significance of national and regional particularities and various historical, cultural and religious backgrounds must be borne in mind, it is the duty of States, regardless of their political, economic

⁹⁹ GC 34 (n 52) 5–6 [21].

¹⁰⁰ [2019] VCC 2142, 30 [158].

¹⁰¹ Special Rapporteur Report A/74/486 (n 14) 5 [6].

and cultural systems, to promote and protect all human rights and fundamental freedoms.¹⁰²

The indivisibility of human rights has now become an ‘official doctrine’ of the UN.¹⁰³

Freedom of expression is unfavourably relegated in state and territory, as well as Commonwealth, anti-vilification legislation. The language of prohibition is uncertain and broadly expressed. Everything depends on the limited domain secured for free speech by narrowly framed exclusions. The exclusions fail to protect freedom of expression in its plenary dimensions. The interpretation placed on anti-vilification laws by commissions, tribunals and courts is driven principally by a statutory text without regard for freedom of expression. At no stage in applying the legislation is the process of justification required by art 19(3) ever undertaken.

During her tenure as President of the AHRC, Professor Gillian Triggs made some important observations of direct relevance to our concerns in this area. She was commenting on the fact that ‘Australia has not implemented the core human rights principles, set out in the [ICCPR], including freedom of speech’ and that ‘[i]n the Australian Constitution, there are only minimal protections provided for basic freedoms — and no explicit protection for the right to freedom of speech’.¹⁰⁴ She observed that ‘Australia’s unique approach to human rights has produced a significant gap in legal protections for some rights, such as the right to freedom of speech ... while comprehensively preserving the right protection against certain forms of discrimination’.¹⁰⁵ This led Professor Triggs to an exhortation to find an appropriate balance between rights, stating,

[b]ut let me be clear. All human rights should be protected. There is no hierarchical order amongst them. Human rights are seldom absolute. Nor are they isolated from one another. The challenge is to find a balance between rights.¹⁰⁶

But then, in an inversion of her stated concern, Professor Triggs continued: ‘It is not appropriate to cherry pick one’s favourite right over another — such as the right to freedom of speech [for which she accepts there is inadequate protection] over the right to anti-discrimination [which is comprehensively protected]’.¹⁰⁷ Her statement seems to suggest that the motive of those who single out freedom of speech for remedial improvement is that this is ‘their favourite right.’ Advocating better protection for freedom of expression in Australia, when it is

¹⁰² *Vienna Declaration and Programme of Action*, UN Doc A/CONF.157/23 (12 July 1993, adopted 25 June 1993), 5 [5] (‘Vienna Declaration’).

¹⁰³ James W Nickel, ‘Rethinking Indivisibility: Towards a Theory Supporting Relations Between Human Rights’, (2008) 30(4) *Human Rights Quarterly* 984, 985.

¹⁰⁴ Gillian Triggs, ‘We Need More Laws, Not Fewer, to Protect Our Freedoms’, *The Guardian Australia*, (online, 22 January 2014) <<https://www.theguardian.com/commentisfree/2014/jan/22/we-need-more-laws-not-less-to-protect-our-freedoms>>.

¹⁰⁵ *Ibid.*

¹⁰⁶ *Ibid.*

¹⁰⁷ *Ibid.*

notoriously deficient, especially compared with the protection given to other ICCPR rights, surely cannot be said to involve any such cherry picking. ‘Trumping’ is another common pejorative description. In the scheme set by ICERD and the ICCPR, there is no so-called cherry-picking or trumping of one right over another if the key texts and principles elaborated by their respective monitoring bodies are heeded.

VI LEGISLATIVE PROPOSALS

In the light of the preceding discussion, we propose certain legislative options. The fundamental aim is to adopt the balancing test required between protection against hate speech and freedom of expression established by ICCPR art 19(3). The options represent proper exercise of the external affairs power conferred by s 51(xxix) of the *Constitution*, as interpreted by the High Court,¹⁰⁸ because of their adherence to the convention standards invoked. They involve Commonwealth legislation that would specifically override the operation of state and territory vilification laws to the extent of any inconsistency by virtue of s 109 of the *Constitution*.

The enactment of overriding Commonwealth legislation is appropriate and necessary. A uniform approach is required to render existing legislation comprehensible. It would produce needed uniformity in an area of law that addresses a category of conduct (speech or expression) that, as discussed, has an inherently trans-jurisdictional character. It will also answer broader concerns of the Human Rights Committee about the inconsistencies in anti-vilification laws that currently exist across different states and territories in Australia.

A Proposal 1: The Minimalist Application of Article 19(3)

A minimalist approach involves the Commonwealth expressly implementing the balancing standard of art 19(3), but in a confined way. The intention is for a Commonwealth law to expressly override the operation of state and territory vilification laws in so far as their application would restrict freedom of expression to an extent contrary to the specific requirements of art 19(3).

Conformity to international standards has been achieved swiftly and simply in this manner in the past. This occurred when the Commonwealth Parliament was faced with the urgency of meeting ICCPR standards under art 17, following Tasmania's refusal to repeal criminal provisions concerning sexual conduct involving consenting adults acting in private. (The background to the change was a complaint brought before the Human Rights Committee in *Toonen v Australia*, a landmark case that successfully challenged Tasmania's laws criminalising

¹⁰⁸ *Commonwealth v Tasmania* (1983) 158 CLR 1; *Richardson v Forestry Commission* (1988) 164 CLR 261; *Victoria v Commonwealth* (1996) 187 CLR 416.

consensual sex between adult males in private on the basis of violating the right to privacy in art 17 of the ICCPR.)¹⁰⁹ By a neat formula in the *Human Rights (Sexual Conduct) Act 1994* (Cth), the Commonwealth gave effect domestically to the concept of ‘arbitrary interference with privacy within the meaning of article 17 of the [ICCPR].’¹¹⁰ Its main effect was to overrule the operation of state and territory law with respect to a narrow aspect of ICCPR art 17, and thereby ended the most obvious aspects of Tasmania’s stand against such reform. It may have been limited in its effect, but it is an example of a faithful, accurate, and economical way of adopting ICCPR demands in an area where it was needed and had to be addressed head-on.

An adaptation of the same legislation to present circumstances might be in the following terms, and is offered here to illustrate available options rather than as a particular model:

No one exercising their freedom of expression, within the meaning of Article 19(2) of the International Covenant on Civil and Political Rights (ICCPR), is to be subject, by or under any law of the Commonwealth, a State or a Territory, to any restriction that is not justified in accordance with the terms of Article 19(3) of the ICCPR.

This approach might be criticised for requiring too much judgement and regard for principles established by the Human Rights Committee. This is also true when determining ‘arbitrariness’ under art 17, but that was no impediment to the 1994 legislation. The proposed text would operate largely as a backstop, to ensure due regard for art 19 in those cases where it is likely that the free speech exclusions from anti-vilification prohibitions do not go far enough. It is modest as a remedial solution. The proposal would have no effect on the application of state and territory vilification laws to instances of hate speech of greater severity.

B Proposal 2: General Effect to the Totality of Article 19(3)

A broader alternative to Proposal 1 is an enactment that gives general effect to the totality of art 19, in substance, for all purposes. It would render inoperative (ie, ‘invalid’ pursuant to s 109 of the Constitution) any restriction on the freedom of expression save strictly in accordance with terms similar to art 19(3). The intention would be to entitle those on whom such restrictions are imposed to adjudication by a competent authority, and a remedy, as required by ICCPR art 2(3). This is no more than what the Human Rights Committee already expects from Australia given its obligation to ‘adopt such laws or other measures as may be necessary to give effect to the [ICCPR] rights’.¹¹¹ Anti-vilification laws of states

¹⁰⁹ *Toonen v Australia*, Communication No. 488/1992, Views of 31 March 1994, CCPR/C/50/D/488/1992.

¹¹⁰ ‘Sexual conduct involving only consenting adults acting in private is not to be subject, by or under any law of the Commonwealth, a State or a Territory, to any arbitrary interference with privacy within the meaning of Article 17 of the International Covenant on Civil and Political Rights’.

¹¹¹ ICCPR (n 6) art 2(2).

and territories would continue to operate subject to such Commonwealth enactment.¹¹²

C Proposal 3: Prohibition on Particular Conduct with Known Article 19(3) Justification

Proposal 3 is intended to apply in combination with whichever of the above Proposals (1 and 2) is adopted. It consists of Commonwealth enactment of hate speech provisions corresponding with the texts of *ICERD* art 4(a) and *ICCPR* art 20(2). This would represent simple implementation enabling Australia to remove its reservations in respect of both.¹¹³ As we hope we have demonstrated, it is clear that the justifications required for the resulting prohibitions satisfy art 19(3) criteria.

The enactments could be accompanied by clarification that it is not intended to exclude or limit the operation of a law of a state or territory that furthers the objects of *ICERD* article 4(a) consistently with Australia's obligations under that provision and is capable of operating concurrently with the Commonwealth law.¹¹⁴ However, to achieve that consistency by upholding freedom of expression under *ICCPR* art 19, Proposals 1 or 2 should be implemented in tandem with Proposal 3.

Within this limited scope Australia would achieve uniformity on the topic of racial and religious hate speech.

D Proposal 4: Non-Legislative Options

Proper regard should also be had to non-legislative options. The Special Rapporteur recently made the point in this way: 'The recognition of steps other than legal prohibitions highlights that prohibition will often not be the least restrictive measure available to States confronting hate speech problems.' Valuable alternatives include public statements by leaders in society countering hate speech, fostering tolerance and intercommunity respect, education and intercultural dialogue, enhanced access to information and ideas counter hateful messages, and the promotion of and training in human rights principles and standards.¹¹⁵ Comparable alternatives are also recommended in the *United Nations Strategy and Plan of Action on Hate Speech*.¹¹⁶

¹¹² Another possible approach would be for the states and territories to enact cooperative legislation. However, the likelihood of achieving consensus across all the jurisdictions is, at this point in time, remote.

¹¹³ Australia's reservation did not reject art 20(2) outright, but rather reserved the right not to introduce any further legislative provisions on the topic: Temperman (n 10) 72–73.

¹¹⁴ See, eg, *Racial Discrimination Act 1975* (Cth) s 6A(1).

¹¹⁵ Special Rapporteur Report A/74/486 (n 14) 9 [18].

¹¹⁶ *Strategy and Plan of Action* (n 1) 4.

VII CONCLUSION

Hate speech is, sadly, increasingly widespread. The social climate in Australia and many other countries is becoming increasingly politicised, polarised and hostile. Social media has become a powerful channel for political and cultural influence, and has fuelled divisiveness, including by demonising individuals and groups. They may be pilloried for the views or beliefs they do not hold, as much as for those they do hold, through the stereotypical ascription of a set of negative attributes. This is the essence of much status-based phobia, but it has also become a common device of persuasion. The phenomenon is not confined to social media. It is also reflected in institutional cultures. People are called out for their vilifying speech (whether it is that or not) by those indulging in serious vilification themselves. In this environment, anti-vilification legislation is at risk of becoming weaponised. This is more likely to occur at the threshold for vilification, by interpreting it downwards, and is more likely to occur the greater the uncertainty that attends anti-vilification legislation. There is an increasing risk that speech is able to be regulated on the basis of its political content.

One legislative aim of all anti-vilification legislation is self-censorship by those who would otherwise indulge in harmful speech. Findings that the legislation is contravened reaffirm community standards and serve as a public warning against others making the same mistake. However, to win and maintain public support, the legislation needs to be predictable. It falls into disrepute where it operates at too low a threshold, or where self-censorship occurs out of uncertainty or fear of reprisal. Australian anti-vilification legislation, in its current state, is confronting. It comprises a multiplicity of statutes, prohibiting speech on different grounds, at different thresholds, for different activities, and by different means (varying by state and territory between criminal and civil, civil only, and criminal only). The legislation becomes especially intimidating when interpreted or applied by competent authorities unpredictably against public figures in the full glare of publicity. There is every reason to anticipate that those who observe this occurring will be cowed into silence.

The discussion in this article leads to recommendations for reform to prevent state and territory anti-vilification legislation being applied contrary to the dictates of ICCPR art 19(3). The fault is not so much in how the existing legislation affects the most serious instances of vilification, but in how it operates at the opposite end of the spectrum, closer to the threshold where uncertainty is greatest. The uncertain and varying thresholds applied across Australian jurisdictions mean that the best way of restoring certainty is to adopt a single standard of exclusion from the prohibitions, as required by art 19(3). Beyond this, the approaches proposed in this article merely aim to implement core convention hate speech obligations. After all, the purpose of international conventions that mandate the prohibition of hate speech is to address the menace it poses to

democratic values, social stability and ultimately peace. As the *Camden Principles*, which were influential in formulating the *Rabat Plan of Action*, observe:

Pluralism and diversity are hallmarks of freedom of expression. Realisation of the right to freedom of expression enables vibrant, multi-faceted public interest debate giving voice to different perspectives and viewpoints. Inequality results in the exclusion of certain voices, undermining this. The right of everyone to be heard, to speak and to participate in political, artistic and social life are, in turn, integral to the attainment and enjoyment of equality. When people are denied public participation and voice, their issues, experiences and concerns are rendered invisible, and they become more vulnerable to bigotry, prejudice and marginalisation.¹¹⁷

Simple conformity of anti-vilification laws with ICCPR standards should not be so hard to achieve, surely? If not, why not?

¹¹⁷ *Camden Principles* (n 53) 3.