The impact of section 18C and other civil anti-vilification laws in Australia

Luke J. McNamara
University of Wollongong, lukem@uow.edu.au

Katharine Gelber
University Of New South Wales

Follow this and additional works at: https://ro.uow.edu.au/lhapapers

Part of the Arts and Humanities Commons, and the Law Commons
The impact of section 18C and other civil anti-vilification laws in Australia

Abstract
This paper reports on the findings of a large scale study of the impact of anti-vilification (or 'hate speech') laws, on public discourse in Australia over more than two decades. Its scope includes, but is not limited to s 18C of the Racial Discrimination Act 1975 (Cth). We investigated the ways in which legislation might have affected public discourse over time. Our task was methodologically challenging, for connecting changes in public discourse to the introduction or enforcement of hate speech laws is fraught with difficulty. We triangulated data from a range of primary and secondary sources, to investigate the relationship between hate speech laws and public discourse over time. Sources include complaints data from, and interviews with, federal and state/territory human rights authorities; tribunal and court decisions; qualitative document analysis of letters to the editor published in newspapers; data from community organisations regarding their members’ experiences; and interviews conducted with members and representatives of target communities.

Keywords
section, 18c, other, anti, impact, civil, laws, vilification, australia

Disciplines
Arts and Humanities | Law

Publication Details

This conference paper is available at Research Online: https://ro.uow.edu.au/lhapapers/2095
The impact of section 18C and other civil anti-vilification laws in Australia

Luke McNamara & Katharine Gelber

Introduction

This paper reports on the findings of a large scale study of the impact of anti-vilification (or ‘hate speech’) laws,1 on public discourse in Australia over more than two decades.2 Its scope includes, but is not limited to s 18C of the Racial Discrimination Act 1975 (Cth). We investigated the ways in which legislation might have affected public discourse over time. Our task was methodologically challenging, for connecting changes in public discourse to the introduction or enforcement of hate speech laws is fraught with difficulty. We triangulated data from a range of primary and secondary sources, to investigate the relationship between hate speech laws and public discourse over time. Sources include complaints data from, and interviews with, federal and state/territory human rights authorities; tribunal and court decisions; qualitative document analysis of letters to the editor published in newspapers; data from community organisations regarding their members’ experiences; and interviews conducted with members and representatives of target communities.

Five claims about the effects of hate speech laws

The chief organising concept for our project was an investigation of five of the most important and cogent claims made about the likely effects of hate speech laws.

The first claim is that hate speech laws provide a remedy to targets of hate speech. Australian laws are sufficiently broad to include both personally targeted vilification directed at an individual or a group, as well as speech that puts into circulation discriminatory views. This reflects the fact that the laws are designed to provide a remedy for both the personal assault on dignity experienced by targets, and the enhanced risks of discrimination and violence that

---

1 We use the terms ‘hate speech’ interchangeably with ‘vilification’ to mean expression that is capable of inciting prejudice towards, or effecting marginalisation of, a person or group of people on a specified ground (adapted from Gelber, K., & Stone, A., (eds) 2007., ‘Hate Speech and Freedom of Speech in Australia’, Federation Press, Sydney: xii). We use it the latter being used in the Australian regulatory framework.

2 We acknowledge funding from the Australian Research Council (DP1096721), and note that ethics approval for this project was granted by the University of Queensland (2011000341). We thank Jess Todhunter, Dave Eden, Sorcha Tormey and Ellyse Fenton for research assistance, and acknowledge the important work undertaken by Cultural and Indigenous Research Centre Australia (CIRCA). We are grateful also for the assistance of the relevant authorities and community organisations from whom we obtained data.
flow from allowing discriminatory stereotypes to circulate publicly. It follows that, in considering whether laws in Australia provide a remedy to the targets of hate speech, we consider two conceptions of ‘targets’. The first are individuals who have been personally targeted, whether face-to-face, or by being named in a statement communicated to the public (e.g., newspaper article, radio program, website). The second are members of a targeted group, whether or not they individually were subjected to, or heard, the conduct in question.

A second core idea is that hate speech laws will, or ought to, have a constructive effect on public discourse by encouraging more respectful speech. Such laws are not designed to silence discussion on controversial topics, but to underpin an obligation to present opinions in a ‘decent and moderate manner’\(^3\). Prior research in Australia has suggested precisely that they are designed to proscribe ‘incivility in the style and content of publication of racist material’\(^4\), or even that, in attempting to regulate for civility, they privilege the ‘racist acts of social elites’\(^5\), although other research has suggested these interpretations are too narrow\(^6\).

The third alleged effect of hate speech laws that we will consider is whether they have an educative or symbolic value. This is the idea that the laws make a statement by government that discourse of a certain type is unacceptable. Jeremy Waldron has described this goal as a publicly expressed commitment to uphold people’s dignity\(^7\). Importantly, this claim is independent of whether hate speech laws are invoked in any particular instance.

The fourth claim is that these benefits can be achieved without producing a ‘chilling effect’ on speech. The fifth is that the risk of creating ‘martyrs’ is outweighed by the potential for authoritative condemnation of hate speech. These claims are rebuttals of two of the primary objections made by opponents of hate speech laws. As Schauer\(^8\) has pointed out, many laws are designed to ‘chill’ in the sense of deterring people from engaging in harmful behaviour. Chilling of this sort is considered to be laudable. Critics of hate speech laws use the term ‘chilling effect’ in a pejorative sense, connoting that individuals might be discouraged from

engaging in legitimate political debate for fear of falling foul of legislation that proscribes hate speech. The risk of creating martyrs has been explained as follows:

... Judicial determinations of guilt or innocence under 'hate speech' laws have social implications that ... can create 'martyrs' of those who would incite discrimination and can claim to have been unjustly silenced by the state ... such offensive expression is given more public attention than it might otherwise have received.

Proponents of hate speech laws claim that neither of these risks represents a compelling argument against creating legal regimes for delineating forms of unacceptable speech, and that they overstate the potentially negative effects of hate speech laws and downplay their benefits.

**A remedy for harms?**

There are two ways in which we construe a ‘remedy’. The first is whether targets are able successfully to lodge complaints for incidents of hate speech and achieve an outcome that ameliorates its effects. The second is whether the laws have contributed to a reduction in the frequency or virulence of hate speech.

A useful starting point in answering the first question is the number of complaints lodged with authorities since the laws were introduced. Our collection of complaints data from all Australian jurisdictions revealed that the number of complaints in any given year is relatively modest. In the decade up to 2010, the total number of complaints nationally fluctuated from a high of 342 to a low of 165 per year. These are relatively modest numbers of complaints, given the size of the Australian population at approximately 20 million, and the extent of anti-vilification laws that cover most jurisdictions and a variety of grounds.

We observed a trend, shortly after new legislation is introduced, to test it out, as evidenced by relatively higher numbers of complaints compared with later periods. For example, the year 2004-05 shows a significant increase in the numbers of complaints compared with the previous year. Nearly half of these complaints were in one jurisdiction – Tasmania – and occurred

---

shortly after the introduction of that state’s anti-vilification laws. This peak is thus explicable as an example of the higher use of a complaints mechanism shortly after its introduction.

It may be that the higher use of the law in the first few years after it is introduced is due to a heightened awareness of the newly-enacted legislation, combined with a desire to test its utility and application. This suggestion was supported by Jeremy Jones, who, as an elected official of the Executive Council of Australian Jewry (ECAJ), has been instrumental in invoking racist hate speech laws to address anti-Semitism. Jones told us in interview that when the laws were first introduced, their organisation looked at, ‘where do people feel most unable to respond as individuals, and where are we getting people saying we have to do something?’13 These cases were pursued and clarification of key aspects of the law’s operation obtained, including the threshold required for an incident to be actionable, that material on the internet was covered by the provisions, and that Holocaust denial was prohibited. Subsequently, the community was able to use those judgments in combatting other incidents:

You have a newspaper that’s published something, you say, ‘look at the rules, look at this judgment’, and people say ... ‘we didn’t know, we didn’t realise, now we do, we don’t want to break the law’.

The judgments were used as a tool of advocacy to convince people not to engage in vilification. This was the case even though less than two per cent of matters are resolved by formal adjudication in a tribunal or court, and therefore produce judgments that are released publicly. Where matters are resolved by confidential conciliation, there is very limited opportunity to use these outcomes for educational purposes. The human rights authorities report on some anonymised case studies in their annual reports, but do not release data that list how many hate speech complaints they have dealt with or what those complaints involved. This contributes to what we discovered in interviews with members of targeted communities: that public awareness of the existence and nature of hate speech laws is uneven and, in some communities, low.

After a ‘peak’ shortly after legislation was introduced, most jurisdictions see a drop-off in the number of complaints over time. For example, in New South Wales, 2009-2010 saw only 22 complaints of racial vilification lodged. There are a number of possible explanations for this drop off. One is that there is less need for the active engagement of the law because the community improves its discourse. This was the view expressed by a former Attorney-General for New South Wales. Commenting on public submissions to a review by that state of its never-

13 Jones, J., (2013) Interview conducted by authors, unpublished.
prosecuted criminal anti-vilification laws, Mr Dowd said the decline in the number of complaints over time indicated that the law was achieving its educative purpose\textsuperscript{14}.

However, there is evidence to contradict this assertion. First, previous research has shown that the majority of hate speech matters terminate before a conciliation is achieved, due in part to some complaints lacking substance, but more usually to procedural barriers including the need to identify and locate the respondent, and the long time that it can take before a complaint reaches conciliation in some jurisdictions\textsuperscript{15}. Second, there is evidence that the incidence of hate speech in the community has remained at concerning levels. Numerous reports from community organisations have pointed to ongoing high levels of verbal abuse suffered by target communities. For example, Jeremy Jones, who has for twenty years maintained a database on incidences of anti-Semitic ‘racist violence’,\textsuperscript{16} recorded a significant increase in verbal harassment from 8 in the year ending September 1990, to 128 in the year ending September 2011\textsuperscript{17}.

There are continuing incidences of prejudicial expressions against Arabs and Muslims. In 1998 a report noted that the 1990 Gulf Crisis had created an atmosphere that was ‘conducive to the ‘scapegoating' of Arab and Muslim people\textsuperscript{18}. A 2004 report on religious diversity noted that while in some areas religious communities cooperated well and inter-faith initiatives were burgeoning, nevertheless the terrorist attacks of September 11, 2001 had ‘triggered an Australia-wide spate’ of abuse, hate mail and assault. Veiled Muslim women were a particular target and reported an inability to venture into public\textsuperscript{19}. These findings were replicated in our interviews with members of Arab and Muslim communities who stated that since the 2001 terrorist attacks, members of the wider community felt that it was acceptable to engage in verbal abuse towards them, in part because political leaders were also doing so.

Finally, reports on the experiences of Indigenous Australians demonstrate that verbal abuse is persistent and ongoing. A 2012 report in Victoria noted that 92 per cent of respondents had experienced being called racist names, or being subjected to racist comments or jokes in the

\textsuperscript{16} Jones uses the definition of ‘racist violence’ contained in the AHREOC’s Racist Violence: Report of the National Inquiry into Racist Violence in Australia (1991: 14): ‘verbal and non-verbal intimidation, harassment and incitement to racial hatred as well as physical violence against people and property’.
previous 12 months\textsuperscript{20}. In our interviews, Indigenous people confirmed that they were routinely subjected to verbal expressions of racism that were disempowering, including children in school. This means that it is unlikely that the decline in the number of formal complaints under the civil hate speech laws over time reflects an improvement in the quality of public discourse or a reduction in incidents of hate speech.

We also conducted interviews with ‘successful’ complainants/litigants and their lawyers\textsuperscript{21}. These showed that the complaints most likely to achieve the remedy sought and advance the wider objective of deterrence have occurred when the complainant is supported by a representative organisation, or has exceptional personal resolve to pursue the matter; and where the person alleged to have engaged in unlawful hate speech is an ‘ordinary’ member of the community, rather than a high profile public or media figure. This is because of the commitment required to pursue a complaint to a successful conclusion, and the likely amenability of the respondent to change their behaviour in a system that relies heavily on voluntary compliance with a conciliated settlement.

Successful deployment of hate speech laws ideally relies on an extraordinary individual, backed by a well-respected organisation that provides credibility, resources and expertise. As Jones observed with reference to the case in Tasmania, their first litigated ‘win’ under federal racial hate speech law,

\ldots we had the advantage of an individual [Jones] who had been dealing with this stuff for twenty or more years at that time, We had a lawyer who is very used to industrial law but there were enough similarities and a barrister who had a lot of experience in defamation \ldots For an average member of the public to use the law [is] extremely difficult.

Jones adds that because he had been documenting anti-Semitic incidents for years, the ECAJ had an evidence base to support informed, strategic decision-making about which matters should be litigated. No other community affected by hate speech in Australia has documented the problem to the same extent.\textsuperscript{22}

\textsuperscript{20} Victorian Health Promotion Foundation (VHPF) (2012) Mental Health Impacts of Racial Discrimination in Victorian Aboriginal Communities: Experiences of Racism Survey – A Summary. Melbourne: VHPF. At 2


\textsuperscript{22} In 2014 a website and Facebook page were launched called the Islamophobia Register Australia: http://www.islamophobia.com.au (Veiszadeh 2014).
We do not suggest that hate speech laws can only be successfully invoked in these circumstances. There is evidence to the contrary.\textsuperscript{23} However, our interviews with litigants and members of targeted communities supported this view.

Keysar Trad’s long-running battle with radio personality Alan Jones provides another example of the heavy burden carried by complainants/litigants under Australian civil hate speech laws. In April 2005, Jones made statements during his Sydney radio broadcasts including calling Lebanese Muslims ‘mongrels’ and ‘vermin’, and saying they ‘hate our country and our heritage’, ‘have no connection to us’, ‘simply rape, pillage and plunder a nation that’s taken them in’, were a ‘national security problem’ who were ‘getting away with cultural murder’, and making women feel unsafe and threatened. It took nearly a decade for Trad to achieve what he set out to achieve: confirmation that Jones comments were unlawful under racial vilification under s 20C of the \textit{Anti-Discrimination Act 1977 (NSW)}\textsuperscript{24}.

Such stories confirm that Australia’s primary model of hate speech regulation places a heavy burden on the targets of hate speech. The legislation can only be invoked in relation to a given incident if a member of the vilified group is willing to step up and take on the arduous, stressful, time-consuming and possibly expensive task of pursuing a remedy on behalf of the wider community. In a sense, the regulatory model \textit{assumes} the existence of such a person in each of the targeted communities. As a result the benefits of the protection of Australian hate speech laws have been unevenly distributed, depending on the ability and willingness of the affected community to pursue hate speech litigation.

Overall, this analysis indicates that civil hate speech laws are providing a remedy, in two senses. The first is that complaints can be lodged and in some cases a favourable outcome obtained. The ability to have a governmental authority validate the message that hate speech breaches the law is important in and of itself, since it provides targeted communities with the knowledge that the law can assist in protecting them from discrimination. The second is in terms of the laws’ educative role. That this educative role includes directly using precedents to dissuade hate speakers is of particular interest, since it would not be able to occur in the absence of hate speech laws. Given the ability of the civil hate speech model to target a wider range of expressive conduct than a purely criminal model would permit, this is particularly


important. It provides direct evidence of the educative role that hate speech laws can play. The remedies are, however, limited as there are persistent, significant levels of hate speech, the burden on complainants in seeing a complaint through can be high, and there is an uneven distribution of the benefits among target communities\(^\text{25}\).

**A modification of speech?**

We now consider other evidence regarding whether an improvement of discourse has occurred. We have already established that hate speech is ongoing. Here, we provide further data to consider whether there has been a reduction in hate speech over time.

We conducted a qualitative document analysis of 6612 letters to the editor published in newspapers in each jurisdiction over the period of the study\(^\text{26}\). There is a difference between language use in the mediated outlets that are newspapers, and language use on the street (explicated above). We view the letters to the editor as a mediated discourse that demonstrates the tension between publishing views of members of the public on the one hand, and remaining within the confines of legally permissible expression on the other.

We found a discernible shift in language used to express prejudice towards Indigenous peoples. In the early 1990s terms such as ‘uncivilised’ and ‘not civilised’ were in prejudicial letters. By the mid-1990s prejudice was primarily conveyed by referring to ‘frivolous title claims’, ‘special laws for Aborigines’, the ‘Aboriginal guilt industry’, and the Stolen Generations ‘myth’. We found no consistent shift in language used to express prejudice towards recent migrants, with expressions including ‘send migrants back where they came from’, ‘ethnic crimes’, ‘noisy minorities’, and descriptions of asylum seekers since circa 2000 including ‘human evil’, ‘illegal immigrants’, ‘terrorists’, ‘uninvited intruders’ and ‘queue-jumpers’.

Overall, there was a modest but significant reduction in the expression of prejudice over time. When the letters are divided into three equal time periods, the proportion of ‘prejudicial’ letters published in 1992-1997 was 33.86 per cent, in 1998-2003 the figure was 29.08 per cent and in 2004-2009 the figure was 28.54 per cent. Of course, our data cannot tell us clearly the extent to which hate speech laws themselves contributed to this reduction in mediated expressions of prejudice and we acknowledge that a myriad of social factors has contributed to this change.

---


Nevertheless, the laws likely played a part in forming the climate within which newspapers are publishing fewer prejudicial letters.

Interviews with members of targeted communities also yielded insights into whether hate speech laws have exerted a positive influence on discourse. Indigenous interviewees tended to be pessimistic, stating that the prevalence of hate speech towards Aboriginal and Torres Strait Islander people over time had remained the same, or increased. One interviewee said,

If you’ve got commentators who are out there with their hate speeches, a lot of it can be dressed up as acceptable speech, when, in actual fact, it’s totally unacceptable. But, somewhere along the way, we’ve kind of been numbed into accepting that it’s okay …

A common theme in the views expressed by interviewees was that hate speech remained a prevalent feature of life, but that its primary targets had changed. For example, a member of the Vietnamese community felt that things had improved (compared to the 1980s and 1990s) for Vietnamese people in Australia, but that racist attention had shifted to Muslims and more recent immigrant communities from Afghanistan and Africa. This view was echoed by Sudanese and Afghan interviewees. A Turkish interviewee said,

I think it shifts from community to community … so it might have been sixty, seventy years [ago] or whatever, the Italians and the Greeks, then the Middle Eastern [and] Turkish people, then it shifted to the Chinese, now to the … African and the Afghani community.

No interviewees thought that hate speech laws had had a profoundly positive influence on the quality of public discourse. However, a number were of the view that the laws had yielded some benefits:

Has legislation had an impact on the level of hate speech? I think it has to a certain extent. It does not mean it’s eliminated it … But people are more conscious and aware of it … it has curtailed some of the utterances that people might hold back … So the legislation has had some role in perhaps reducing or minimising that harm.

An educative and symbolic effect?

Is there evidence from our study that Australian hate speech laws have had an educative effect on the public, or provide a symbol of support for targeted communities? We have already concluded that there have been two ways in which the laws play an educative role. The first is

the direct and conscious use of prior judgments in community advocacy and as a device to curb ongoing vilification by telling the perpetrators that the court has stated that what they are doing is unlawful. A second, albeit less direct and harder to quantify, educative effect has been evidenced by the reduction in the proportion of prejudicial letters published in newspapers. Combined with the evidence of knowledge of the existence (if not the definition) of hate speech laws among letter writers, it is possible that the existence of hate speech laws has played a role in educating them in how to avoid confrontation with the laws, even if they still wish to express prejudice. However, it is also possible that even successful hate speech litigation can communicate messages that are at odds with the laws’ educational goals. This point was illustrated by the public discourse that emerged in the aftermath of the Federal Court of Australia’s decision that newspaper columnist, Andrew Bolt, had engaged in unlawful hate speech by suggesting that a number of fair-skinned Aboriginal people had deliberately chosen an Indigenous identity over others that were more logically available to them, and that they had done so for personal gain28.

Importantly, in interviews many community members and representatives, when asked if they thought hate speech laws were important, expressed overwhelming support for their retention. There was a strong sense that the laws could make a positive contribution outside their formal utilisation. The overwhelming view was that the laws were useful as a statement in support of vulnerable communities. Interviewees described it as important simply to ‘know they’re there’ and that they set a standard for what’s ‘not acceptable’. It follows that the legal form and parameters of hate speech laws may be less important than the fact of their existence. The Australian experience with civil hate speech laws suggests that a decision not to rely on the criminal law should not automatically be interpreted as a ‘weak’ regulatory response, but rather as a potentially useful way of setting a standard for public debate.

A ‘chilling effect’ or the creation of martyrs?

Our analysis of letters to the editor revealed little evidence that public discourse has been diminished over the past 25 years. Robust debates have been had on a broad range of issues including the land rights of Indigenous Australians, same-sex marriage, and the treatment of asylum-seekers. Our analysis revealed the continued expression of prejudice over time. The fact that we detected a shift away from more intemperate styles of language cannot be said to support the chilling effect claim. At the heart of this claim is a concern about the silencing of views and opinion. Yet at the same time that Andrew Bolt claimed he was being ‘silenced’ by hate speech laws, he was able to disseminate his views widely through prominent media

attention\textsuperscript{29}. Therefore, although the distinction may be contentious, we distinguish between desirable and undesirable effects. Hate speech laws are designed to influence the terms in which individuals express their views in public (desirable), however they are not designed to make certain topics 'off limits' (undesirable). Our research suggests that the risk of a chilling effect has not been substantiated. Australians are willing to express robust views on a broad range of policy issues.

The story of Bolt’s encounter with Australia’s national racial hatred law \textit{does} lend some support to the claim that hate speech laws can produce martyrs. After Bolt was found to have breached s 18C of the \textit{Racial Discrimination Act} 1975 (Cth), an orchestrated reconstruction of the decision dominated media discourse in which Bolt served as a representative victim for a wider class of opinion-holders on issues of Aboriginal identity, hate speech laws as incursions into free speech, and the vulnerability of free speech. These events confirm that the invocation of hate speech laws can have unintended effects that subvert rather than promote their underlying values\textsuperscript{30}.

Yet a sense of proportion is required here. No other case in over two decades of civil litigation has triggered a comparable martyr effect. Recalcitrant Holocaust denier Frederick Toben attempted to adopt a martyr position when he was found to have breached the same federal racial hatred law years earlier.\textsuperscript{31} His refusal to abide by orders of the Federal Court to remove Holocaust denial material from his website resulted in 24 contempt of court findings and, ultimately, a three month jail term for contempt of court\textsuperscript{32}. However, in public discourse this attempt served to consolidate his infamy and status as a powerful illustration of precisely why hate speech laws were enacted in the first place\textsuperscript{33}. Two distinctive features of Australia’s hate speech laws are noteworthy here. First, given that most transgressions of the law are addressed in confidential conciliation, with less than 2 per cent resulting in court or tribunal decisions that enter the public domain, opportunities for martyrdom are rare. Second, because the laws rely overwhelmingly on \textit{civil} remedies, they tend not to produce the criminal sanctions on which the claimed martyr effect is based. The Bolt controversy does not justify a general

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{31} Jones \textit{v. Toben} (2003) 129 FCR 515.
\end{itemize}
\end{footnotesize}
conclusion that hate speech laws necessarily produce a counterproductive martyr effect, as it was an atypical event in the history of civil hate speech laws in Australia.

Conclusions

Our project speaks both to the utility and the inefficacy of the dominant legal model for addressing hate speech in Australia. Anti-vilification laws like s 18C of the Racial Discrimination Act 1975 (Cth) do provide targeted communities with the opportunity to lodge complaints with a human rights authority, in a process that reassures them that the law can assist them, and reminds them that the polity has enacted provisions that enable them to seek redress for hate speech. Further, the laws have educative functions – both direct and indirect – and symbolic importance. On the other hand, we found ongoing and significant levels of hate speech, a regulatory model that relies on individuals who are willing and able to bear the burden of enforcement, and an uneven distribution of benefits among targeted communities.

Despite these mixed results, targeted communities expressed overwhelming support for the value and retention of the laws, as a symbol of their protection and the government’s opposition to racist vilification and discrimination. Indeed, although they are often demonised by strident opponents, it would appear that hate speech laws have become an accepted part of the Australian landscape. A 2014 opinion poll showed 88 per cent of the public supporting the retention of s 18C of the Racial Discrimination Act 1975 (Cth) in its current form. This shows that 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. This gives these laws a normative influence, and provides participants in public debate with a language they can employ to condemn hate speech when it occurs.