THE ONLINE HATE PREVENTION INSTITUTE
Empowering communities, organisations and agencies in the fight against hate.

LA LSIC - AVP INQUIRY
SUBMISSION NO. 38
RECEIVED 17 JANUARY 2020

ONLINE HATE PREVENTION INSTITUTE SUBMISSION TO THE

INQUIRY INTO ANTI-VILIFICATION PROTECTIONS

LEGAL AND SOCIAL ISSUES COMMITTEE LEGISLATIVE ASSEMBLY, PARLIAMENT OF VICTORIA

17 January 2020

PREAMBLE

We thank the Committee for the opportunity to make this submission to the inquiry. Much of today's vilification and most importantly, serious vilification, is occurring online or contains online elements. As the only Harm Prevention Charity in Australia dedicated to protecting people by reducing such harmful vilification, we have a strong interest in this inquiry and in the ways the Parliament can better protect targeted sections of the Victorian community from vilification in general and online vilification in particular.

The inquiry is an important step to assess and plan for the ever emerging and shifting terrain of online vilification and hate. Providing protection and prevention from harm for citizens in this terrain is complex and the Online Hate Prevention Institute (OHPI) has been both a pioneer and now seasoned voice in this space.

The advice provided in this submission is founded upon a recognition of the seriousness that high-impact acts of vilification have on individuals and communities. However, the advice is cognisant of the realities of the scale and limitations of online platforms, legal jurisdictions, effectiveness and impact of legal and regulatory positions. We are also conscious of the need to strike an appropriate balance between countering vilification and respecting individual liberties and freedoms of expression. The recommendations in this report reflect those concerns but also provide concrete capacities to protect and foster Victorian communities through remedies to tackle online vilification and hate.

OHPI is unique as a charity with specialised and proven expertise, methodologies and software tools that give us the capacity to identify, categorise and remove instances of online hate. We have had thousands of offensive and vilifying posts removed across major online platforms.

We have worked with many parts of the Victorian community and Government. Our team has worked with Victoria Police, the Office of Multiculturalism Affairs and Citizenship, Muslim, Jewish and Christian communities across Victoria, Imams, Rabbis, the Aboriginal Communities and many others. We have addressed issues impacting all of Victoria as well as issues impacting specific places like the Melbourne CBD and Bendigo. Our advice has been provided to Victorian, national and international bodies including the UN and UNESCO.

We are active in 'hands-on' action documenting, countering and building understanding about online hate and how to address it. We often partner with communities to empower and defend them against online hate. To this end we are regularly communicating with peak organisations and civil society organisations in Australia and internationally. We also have effective direct channels of communication with technology companies such as Facebook, Google, Twitter and YouTube and work with them to improve the systems that respond to hate and to achieve rapid results on urgent situations. This hands-on approach provides a unique perspective on the granularity of online behaviours and the tactics deployed by perpetrators of online hate.

This report has been produced as a formal submission from the OHPI to the Legal and Social Issues Committee of the Legislative Assembly of the Parliament of Victoria. We hope what we have learned over the last 8 years as specialists in this field of online hate can be of assistance to the inquiry and the Parliament.

Dr Andre Oboler, Mr Mark Civitella and Nasya Bahfen

17 January 2020

ABOUT THE AUTHORS

DR ANDRE OBOLER

Dr Andre Oboler is the CEO and Managing Director of the Online Hate Prevention Institute. He serves as a Vice Chair of the MGA Board of the IEEE Computer Society, a Member of the Global Public Policy Committee of the IEEE, an Executive Member of the Jewish Community Council of Victoria, A Victorian Councillor on the Executive Council of Australian Jewry and as an expert member of the Australian Government's Delegation to the International Holocaust Remembrance Alliance.

Andre was formerly a Senior Lecturer in Cyber Security at the La Trobe Law School, co-chair of the Online Antisemitism working group of the Global Forum to Combat Antisemitism, an expert member of the Inter-Parliamentary Coalition to Combatting Antisemitism and served for two terms with the board of the Quality Assurance Agency – the UK's higher education regulator. His research interests include online regulation and hate speech and the impacts of technology on society. He is a co-author of the book "Cyber Racism and Community Resilience: Strategies for Combating Online Race Hate".

Andre holds a PhD in Computer Science from Lancaster University, Honours in Bachelor of Computer Science and an LLM(Juris Doctor) from Monash University. He is a Senior Member of the IEEE, a Graduate Member of the Australian Institute of Company Directors and a Member of the Victorian Society of Computers and the Law.

MARK CIVITELLA

Mark Civitella is a Director of the Online Hate Prevention Institute. He is a Lecturer in Strategic Communication at La Trobe University and works as a strategic communication consultant. He is a Fellow of the Public Relations Institute of Australia.

Mark has expertise in countering violent extremism and political communications as well as issue and crisis management. He has consulted widely and worked with Monash University's Global Terrorism Research Centre, Victorian Multicultural Commission, Victorian Imams Network, and trained religious and cultural leaders.

Mark has been a director of a number of communication and research companies. He holds a B.A. (Monash University); Diploma in General and Comparative Literature (Monash University); Graduate Diploma in Public Relations (RMIT University) and is currently Doctoral Candidate at La Trobe University.

DR NASYA BAHFEN

Dr Nasya Bahfen is a Director of the Online Hate Prevention Institute. She is a Senior Lecturer in Journalism at La Trobe University and also serves as an AFL Multicultural Community Ambassador. Nasya is a former journalist and producer for ABC Radio Australia, ABC Radio National, and SBS in radio and online.

Her research interests include cyber-racism. She was a researcher on the ARC funded Cyber-Racism and Community Resilience project and she is co-author of the book "Cyber Racism and Community Resilience: Strategies for Combating Online Race Hate".

Nasya holds a Bachelor of Journalism from the Royal Melbourne Institute of Technology (RMIT), Honours in Media from La Trobe University, and a PhD in Communications from the University of Technology Sydney.

AUTHORISATION FOR SUBMISSION

This report has been produced by the Online Hate Prevention Institute and approved by the Board of Directors of the Online Hate Preventions Institute for submission to the *Inquiry into Anti-Vilification Protections* of the Victorian Parliament. It draws in part upon an expanding body of material and reports produced by the Online Hate Prevention Institute in its assistance to government, organisations and regulators across Australia and internationally.

[redacted]

Martin Splitter
Chairman, Online Hate Prevention Institute

17 January 2020

RESPONSES TO THE TERMS OF REFERENCE

1) THE EFFECTIVENESS OF THE OPERATION OF THE RACIAL AND RELIGIOUS TOLERANCE ACT 2001 (THE ACT) IN DELIVERING UPON ITS PURPOSES

A) Effectiveness as a message on values

The Act is effective in sending a message that racial and religious vilification go against Victorian values and are not acceptable in Victoria.

It has the added benefit of making it clear that religious vilification is to be treated the same as racial vilification, a much more widely recognised and accepted social harm.

This is particularly effective in countering a range of hate filled memes and narratives we see online which seek to normalise vilification against the Muslim community. This argument has also been used against Jews with the (incorrect) narrative that Jews are a religion not a race and therefore, they argue, antisemitism should be acceptable. We have recently started to see such hate targeting Christians as well. Examples of these phenomena are presented in Appendix 1.

2) THE SUCCESS OR OTHERWISE OF ENFORCEMENT OF THE ACT, AND THE APPROPRIATENESS OF SANCTIONS IN DELIVERING UPON THE ACT'S PURPOSES;

Criminal Provisions

The serious vilification provisions in the Act are effective when they are used, but they are seldomly used. We believe few frontline police have sufficient understanding of the law to recognise when an offence has occurred. As a result, we believe many cases of serious vilification are not taken forward to prosecution.

We also believe this problem is particularly true with respect to online examples of serious vilification. An online message may not make it clear who it is from, if they are within the jurisdiction, if the threat should be taken seriously, etc. Front line police may too readily dismiss such cases as being “online” and therefore “not real”, while the impact on the victim may be just as great as a face to face encounter. There is an attitude among many online who believe they are immune to real world consequences and the law will not be applied to their serious vilification online. This is only exacerbated in online forums which allow people to post anonymously. In our December 2019 report “Hate and Violent Extremism from an Online Subculture”,[1] we note how the /pol/ community found on places like 4chan and 8chan have been responsible for four deadly terrorist attacks in 2019. Figure 7 in Appendix 1 gives an example of material vilifying Christians and inciting violence against Christians, Muslims and Jews. This could be dismissed as nothing more than talk, except such vilification in this particular community has led to multiple deadly terrorist attacks. Such a threat cannot be ignored. While we don't know the number of Victorians in this community, we know that Australia as a whole is heavily represented and makes up about 5% of the total 4chan audience and in absolute terms Australia is the 4th largest source of traffic to 4Chan.

[1] https://ohpi.org.au/hate-and-violent-extremism-from-an-online-subculture-the-yom-kippur-terrorist-attack-in-halle-germany/

We believe the numbers for 8chan, where the Christchurch attack was announced and which inspired that attack, would be similar.

We note there is a growing exception in respect to online videos, often taken by the person engaging in the offence, which are rightly seen as evidence. There have been a number of successful prosecutions under the law in this regard, most notably of the United Patriots Front. These cases reflect positively on the police, the justice system and the laws.

Civil Provisions

Data from the Victorian Equal Opportunities and Human Rights Commission noted a steep rise in complaints related to race when comparing data from 2016-2017 with data from 2017-2018.[2] Formal complaints rose from 77 to 136 over this period (76%) while the number of reports to the inquiry line rose from 470 to 630 (34%).[3] The most obvious point is that only 16.4% in the first year and 21.6% in the second year converted from inquiries into formal complaints[4]. Given the vast majority of cases of vilification are unlikely to result in even an inquiry, the law here has only limited effect. The burden on the community to take a complaint forward is high, and the result is of little benefit outside of areas like commerce and employment where a financial settlement may address at least some for impact of the discrimination. This is reflected in the VEOHRC data where 88% of the formal complaints fell into these two categories.[5]

Appropriateness of Sanctions

The Online Hate Prevention Institute has long advocated for a lower threshold state enforced penalty, such as a fine or civil penalty, which could be used to address more minor issues of vilification – including when it occurs online. A system which allows multiple fines to lead to an increase in the amount of the fine or potentially more serious consequences such as community service orders would make this more of a deterrent. At present we have greater deterrents for poor parking than for vilification which negatively impacts the fabric of the community.

In the case of serious vilification, while larger penalties are available, the fact that the first use of the Act for serious religious vilification was shortly followed by a repeat offence in very similar terms by one of those who were previously convicted, suggests the barriers to using the law are very high while the penalty is low, making it an ineffective deterrent. Rather than increasing the penalty, we would suggest reducing the barriers to using the law so it could be applied more often and to lower threshold cases. The cases tried so far should be at the upper end of a continuum, subject to higher sanctions, not the lower end where they rest at present.

If a civil penalty or fine was introduced for vilification that fell short of a criminal act, the criminal provisions could be adjusted so that a history of such fines or civil penalties could be taken into account during sentencing.

3) INTERACTION BETWEEN THE ACT AND OTHER STATE AND COMMONWEALTH LEGISLATION;

[2] https://www.humanrightscommission.vic.gov.au/home/news-and-events/commission-news/item/1731-new-figures-show-jump-in-race-discrimination-in-victoria

[3] Ibid.

[4] Ibid.

[5] Ibid.

As our area of focus is online vilification we note that section 474.17 of the Commonwealth Criminal Code covers "using a carriage service to menace, harass or cause offence". The provision does not require the victim to belong to a protected group. It is broad enough to cover both vilification (without a threat of physical harm) and serious vilification (with a threat of physical harm). The penalty is three years imprisonment. This is six times greater than the 6 month penalty for serious racial and religious vilification under the Racial and Religious Tolerance Act.

The stalking provisions in section 21A of the Crimes Act 1958 (Vic) can provide an alternative criminal provision to vilification targeted at an individual. Section 21A(2)(ba) includes the definition of stalking as "publishing on the Internet or by an e-mail or other electronic communication to any person a statement or other material (i) relating to the victim or any other person" while 1A(2)(da) covers "making threats to the victim". The stalking provisions carry a penalty of 10 years imprisonment, much higher than serious vilification.

Both the above laws can cover online vilification targeting an individual. They are not effective against vilification of a group, such as the examples in Appendix 1. Section 18C of the *Racial Discrimination Act 1976* (Cth) is effective against group-based vilification, but only where the vilification occurs on the basis of race. The Victorian Act therefore has greater cover with its religious vilification provisions. Additionally, the Commonwealth's Racial Discrimination Act has no criminal provisions.

4) COMPARISONS IN THE OPERATION OF THE VICTORIAN ACT WITH LEGISLATION IN OTHER JURISDICTIONS;

Victoria provided national leadership when the act was introduced. Until fairly recently the religious vilification provisions were seen as largely symbolic, but welcome nevertheless. Increasingly other Australian states are following. In Victoria these provisions are being put to work.

We draw attention to the *Additional Protocol to the Convention on Cybercrime, concerning the criminalisation of acts of a racist and xenophobic nature committed through computer systems.*[6] This international convention, which 32 countries have ratified,[7] commits states to pass legislation to criminalise "distributing, or otherwise making available, racist and xenophobic material to the public through a computer system",[8] as well as criminalising the making of threats against a person or group on the basis of "race, colour, descent or national or ethnic origin as well as religion if used as a pretext for any of these factors".[9] While the *Additional Protocol* makes mention of religion, it is limited to cases when religion is used as a proxy. This would make religious vilification against a Muslim person from a predominantly Muslim country likely to be protected, while a local convert to Islam would be unlikely to be protected.

We believe the Victorian definition to protect religion, namely making it "on the ground of the religious belief or activity of another person or class of persons" is a better approach to covering religion. At the same time we believe the move to criminalise the distribution or making available online material which vilified a protected group (without needing the added element of a threat of harm) is a better approach as it shifts the burden of

[6] *Additional Protocol to the Convention on cybercrime, concerning the criminalisation of acts of a racist and xenophobic nature committed through computer systems*, opened for signature 28 January 2003 (entered into force 1 March 2006) ('Additional Protocol'). Online at https://rm.coe.int/168008160f, accessed 12 December 2019.

[7] https://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/189/signatures?p_auth=cBbjeslP

[8] *Additional Protocol* Article 3.

[9] Ibid Article 4.

taking the matter forward off the shoulders of the victim. Such criminalisation should we believe be handled as a summary offence with a fine of around 5 penalty units. This is a very low threshold and would put it on a par with the offence of flying a kite in the park to someone's annoyance.[10] That would be a step up from where it is now. At the same time, it could create a record leading to a conviction under an additional offence for more serious or repeat offending.

We also note Germany's *Network Enforcement Law* passed in 2017 which requires platforms to remove content that is manifestly unlawful in Germany within 24 hours and other unlawful content (which may require further analysis) within 7 days. There are provisions for a fine of up to €50 million for non-compliance.[11] This is a different approach as it focuses not on the poster of the vilification but on the social media platforms posted in and whom the Germany Government maintains has an obligation to remove it.

5) THE ROLE OF STATE LEGISLATION IN ADDRESSING ONLINE VILIFICATION.

We have discussed this question in depth in Section 1.2 ("Internet Regulation") of our report on "Hate and Violent Extremism from an Online Subculture". The section is reproduced in Appendix 2. We note a number of doctrinal principles covered at the end of the section which may be useful in formulating legislation in this space.

Some specific areas where we recommend state legislation to address online vilification:

Rapid verification of jurisdiction

Jurisdiction is one of the challenges that complicates online regulation. It is particularly acute with online platforms which store and then pass on a user's communications.

To address this, we believe it is necessary for legislation to require platforms to provide a tool to verify if a user is within the State's jurisdiction. Such a tool should work in real time giving an immediate yes or no to the question of jurisdiction. We note it may be necessary to know whether a user is in Victorian as well as whether they are in Australian jurisdiction.

Serious offences

Serious vilification, involving a threat of harm to people or property, should be handled as a criminal matter. The current law does this, but law enforcement may not have ready access to the information they need in order to effectively access a report. The barriers to obtaining that information may see the matter dropped. The rapid verification requirement above may increase the take up of cases of serious vilification by police.

We have recommended that "Serious hate speech, that which makes threats of violence or incites either violence or hatred, should be immediately reported to authorities."[12] This would be an obligation on platforms who become aware of such content either through algorithms or in response to user reporting.

[10] *Summary Offences Act 1966* (Vic) s 4(d)(i).

[11] AFP, 2017. "Germany imposes €50 million fines on social media firms that don't delete hate speech", *The Local.de* (30 June), at https://www.thelocal.de/20170630/germany-imposes, accessed 12 December 2019.

[12] Part of recommendation 20, page 57 https://ohpi.org.au/hate-and-violent-extremism-from-an-online-subculture-the-yom-kippur-terrorist-attack-in-halle-germany/

Vilification

As discussed above, we believe a civil penalty or summary offence approach for vilification that does not involve a threat of violence would be an improvement over the current system. This would allow the burden to be shifted from the victim to the state. We believe these provisions should apply to online content (and this intent should be made clear), but the law should be drafted so it can also apply to other forms of vilification such as may occur in public places or on public transport. This would meet the doctrinal *principle of generality*.

We note that the state would still only take action if a person within the jurisdiction made a report. It should not be the role of the state to go out looking for every instance of vilification. Instead, the state may (like Germany) require platforms to remove content which breaches the law and seek to fine platforms for failing to comply to a reasonable standard.

Repeat offenders

We believe online platforms have the first responsibility to moderate their content. The penalties the online platforms can apply are, however, limited and our work shows them to be ineffective in the face of deliberate offenders. A range of people, including in Victoria, use social media as a means of vilifying others and some have turned it into a source of income by attracting large online audiences and revenue from that viewership as well as merchandise sales.

We believe state legislation should put platforms under an obligation to report to state authorities any user who repeatedly engaged in vilification and is not responsive to the platforms efforts to prevent such behaviour. Practically this means adding reporting to state authorities as a step in the escalation of penalties the platform can apply. Such reporting should not be a decision, but a requirement when a certain threshold of incidents of abusive behaviour is reached. At this point state authorities should investigate and press charges if appropriate.

Legislation would be needed to require at least major platforms to notify authorities when the threshold is reached and requiring them to provide current and historical data on the behaviour that led up to the mandatory referral.

We have previously recommended that: "Other forms of hate speech [outside of threats of violence] should be removed by the platform, but a log of the incident including the user's account and IP address should be recorded. Users should be informed when a platform takes action against them and should be warned that repeated breaches could lead to a report being made to authorities. Where platform sanctions prove ineffective at altering behaviour, the history of breaches and IP address of the user should be referred to authorities."[13]

Platform obligations

Major platforms should be under an obligation to prevent the display of material which is unlawful in Victoria to users of their platform who are connecting from Victoria. We see no reason why the same standard and time frames could not apply here as applies in Germany.

Our work demonstrates there is currently a much higher level of vilification content exposed to Victorians than would be acceptable in Germany. Our report into antisemitism gathered over 2000 items of antisemitic hate

[13] Part of recommendation 20, page 57 https://ohpi.org.au/hate-and-violent-extremism-from-an-online-subculture-the-yom-kippur-terrorist-attack-in-halle-germany/

speech in three months, all of them visible in Victoria.[14] Our project "Spotlight on Anti-Muslim Internet Hate", partially funded by the Victorian Government, reported on over 1,000 items of Islamophobia gathered over three months.[15] OHPI believe the problem has got significantly worse since these studied were conducted. OHPI has the tools to carry out such work regularly to monitor the situation if the resources to support that work were available.

6) THE EFFECTIVENESS OF CURRENT APPROACHES TO LAW ENFORCEMENT IN ADDRESSING ONLINE OFFENDING.

We believe the approach to law enforcement to address online offending could be more effective.

The current approach to vilification means it is not a criminal offence. The public feel it should be a police matter and seek to report activity to the police, but beyond taking a statement little eventuates as there is no underlying criminal offence for police to pursue. This runs against public expectations.

The current approaches to serious vilification set a very high bar before law enforcement pursue a matter. If it is pursued, the offences themselves have relatively low maximum penalties. There is a mis-match between the barriers, both legal and cultural, to pursuing a matter and the ultimate result in the event of a successful conviction. This makes the current approach less effective in practice than it could be, even though it continues to have value (largely) symbolically.

We believe the effectiveness could be improved with a number of changes such as:

Increasing the penalties for serious vilification offences. Offences involving a threat of physical harm to people or property should be seen as a similar level of crime to stalking and should have a similar 10 years imprisonment penalty. These offences should be used sparingly (as they are now) but when they are used the result would then have greater deterrent effect. This should apply to online as well as offline offending.

There should be lower level criminal offences and / or a civil penalty provision available for vilification. One approach might be a summary offence with a five penalty-unit fine. This should apply to both online and offline offending and be pursued by police. Another approach would be a civil penalty provision which could be enforced by a body like the Victorian Equal Opportunities and Human Rights Commission, again for both online and offline offending. Either approach would remove the burden of pursuing the matter from the victim. Both together could allow a system of escalation that initially avoids a police response, but allows a referral to police for repeat offenders. Non-compliance with a civil penalty order could allow for the order to be revoked and a summary offence instituted in its place. This could occur instead of having to go to court to have the civil penalty made enforceable.

Outside of commerce and employment, where the vilification may have caused significant economic loss, damages may not be able to to be corrected for the harm the vilification caused. What victims want, particular in the case of online vilification, is evidence the state takes the matter seriously, evidence the perpetrator has been told they were in the wrong, and evidence there have been some consequence for the perpetrator. In most cases victims

[14] Andre Oboler, 2016. *Measuring the Hate: The State of Antisemitism in Social Media*. Melbourne, Australia: Online Hate prevention Institute. Online at https://ohpi.org.au/measuring-antisemitism/, accessed 12 December 2019.

[15] Andre Oboler, 2015. *Spotlight on Anti-Muslim Internet Hate Report*. Melbourne, Australia: Online Hate prevention Institute. Online at https://ohpi.org.au/anti-muslim-hate-interim-report/, accessed 12 December 2019.

would be happy with a largely symbolic consequence so a low level fine or civil penalty (retained by the state) will be sufficient.

Frontline law enforcement should be better trained in relation to online vilification. Many frontline police do not have a sufficient understanding of the online context to be comfortable taking complaints about online vilification and then pursuing them. Beyond this basic knowledge, given the volume of vilification, it needs to become possible for frontline law enforcement to take these cases further without the need for specialist support. This means additional training, as well as providing them with online tools and a methodology to follow in such cases. The few organisations internationally who have similar expertise to the Online Hate Prevention Institute often provide such training under contract to the police in their jurisdiction.

Protecting fundamental rights and freedom of speech. At present the bar to police taking action is so high that there is little chance of a negative impact on fundamental rights. As the response to vilification, and particularly online vilification, is lowered, there needs to be a greater focus on ensuring fundamental rights are still respected. Protections should be put in place to ensure offences required a profound impact, one that goes beyond the level of a mere slight. Causing fear, humiliation or a sense of isolation from the wider community should be considered a sufficient level of harm to the individual and the social fabric of the community for an offence to occur.

Noting the application to online activity. While we generally prefer that general laws against vilification are used online rather than having special online only laws, we strongly support the approach of the current legislation which includes a note stating the provisions also apply online. This makes it clear that online activity is not outside the operation of the law, while still ensuring conformity between online and offline regulation.

7) ANY EVIDENCE OF INCREASING VILIFICATION AND HATE CONDUCT IN VICTORIA;

Over the last 8 years we have monitored this situation, sharing examples in our reports and briefings. We have been noting a steep rise in the volume of online hate, but also in the tendency for this content to directly incite violence. The incitement is becoming more overt and people are becoming more willing to do so openly under their real names rather than using pseudonyms. We are particularly disappointed to see a number of Victorians turning themselves into internet celebrities and making money out of their promotion of hate.

As we discussed with the BBC in 2015, that year saw “a greater normalisation of hate speech in society than in previous years”.[16] This meant that, "Where previously a person might make a vague negative allusion to race, religion, gender or sexuality, by the end of 2015 the comments on social media were blatant and overt."[17]

In late 2015 / early 2016 we released empirical studies into antisemitism[18] and Islamophobia[19] documenting thousands of examples of this hate and slow or absent responses by the social media companies. With the work being done from Victoria, all the examples were visible here and therefore have an impact on the Victorian community. A number of the examples are specifically Victorian.

[16] https://www.bbc.com/news/blogs-trending-35111707

[17] Ibid.

[18] https://ohpi.org.au/measuring-antisemitism/

[19] https://ohpi.org.au/anti-muslim-hate-interim-report/

Our work looked at Reclaim Australia, the United Patriots Front,[20] and Antipodean Resistance – neo-Nazi youth group who places racist posters and stickers around Melbourne.[21]

Following the "Unite the Right" rally in Charlottesville (USA) in August 2017 which led to the murder of a counter protestor, social media companies clamped down on the far right. In response we saw far right leaders in Victoria being more careful with what they said or making allusions to what they couldn't say. We described this in the media as "walking up to the line spitting over it", while creating a space online where their supporters could reply to their content with more extreme vilification. The moderation of the main content by social media platforms appears to be far better today than their moderation of the comments and replies.

In a series of reports into mass casualty events in Melbourne and overseas such as the Bourke Street attacks (2017[22] and 2018[23]) and the Flinders street attack (2017)[24] we monitored the posts made by major Victorian far right figures and the replies from their supporters. The level of hate had grown so much that it was impossible to read the comments on a single post as quickly as they were coming in. Many of them incited violence. We saw a similar trend continue into 2019 with former Senator Fraser Anning and his visit to Victoria.[25]

Most recently we have shown that Australia is (in absolute numbers) the 4th largest source of traffic on 4chan, the online forum that created /pol/ and online community which now exists on multiple platforms and whose culture of hate is behind the Christchurch terrorist attack and three other deadly attacks in 2019. While we suspect Victoria contributes less to these numbers than some other states, there will still be a significant Victorian element and we are not immune to the risk that the vilification in the /pol/ community might manifest in local extremism.

We have world class tools to conduct further empirical research and real time analysis on the problem, but unfortunately a shortage of funding limits the scope of this important work. This is issue was raised some years ago in the Victorian Parliament and received some interim support at the time, but the problem really needs on-going monitoring which can only be enabled through a partnership with government.

8) POSSIBLE EXTENSION OF PROTECTIONS OR EXPANSION OF PROTECTION TO CLASSES OF PEOPLE NOT CURRENTLY PROTECTED UNDER THE EXISTING ACT;

We believe vilification laws should be consistent in the protection it offers individuals and groups who fall within a protected class. This is to say that different groups should have the same protections as each other, but also that vilification should be treated the same regardless of the medium in which it occurs. Focusing in the online space we seek to address harm that targets any individual or group. We have covered group-based vilification including antisemitism, xenophobia, racism against Indigenous Australians, Islamophobia, homophobia, transphobia, misogyny, racism against a range of other communities and religious vilification against a range of other religions. We believe these groups are particular targets and deserve particular protections. We do not suggest these are the

[20] https://ohpi.org.au/some-upf-supporters-engage-in-extreme-misogyny/; https://ohpi.org.au/the-bendigo-rallies/; https://ohpi.org.au/far-right-harassment-of-senator-sam-dastyari/; https://ohpi.org.au/racists-at-st-kilda-beach/

[21] https://ohpi.org.au/nazi-groups-poster-campaign-melbourne/

[22] A substantial report has been shared with police and is available on request. It has not be published publicly.

[23] https://ohpi.org.au/bourke-street-attack-november-2018/

[24] https://ohpi.org.au/car-attack-in-flinders-street-melbourne/

[25] https://ohpi.org.au/racists-at-st-kilda-beach/; https://ohpi.org.au/senator-annings-hate-machine/

only types of hate that should be specifically addressed, these are just the ones we have seen and responded to as particularly prevalent online.

We note that we have also covered vilification of ANZAC veterans timed for ANZAC Day, serious vilification targeting politicians, we looked at and rejected claims of wide spread online hate against bicycle riders (though the situation on this may have changed since then), and we have dealt with serious trolling of the families of a number of recently deceased children – some members of minority groups mentioned above and others not.

In reality any group that can be identified can be targeted for group-based vilification, but we believe there is value in giving a specific protection to those groups that are particularly vulnerable or are regular targets of vilification. Such protection sends a message that these groups are welcome as part of the community and the community wishes to protect them. It pushes back against those sending a narrative that they are an acceptable target.

Affected communities are best placed to inform government of the impact of vilification on people in their communities and of the impact on the community as a whole. Social media and other online content is a key driver of vilification both directly and indirectly but the impacted communities do not currently have the expertise to effectively tackle this problem. The Online Hate Prevention Institute is recognised as a world leader in this space by organisations like UNESCO and there are in fact very few organisations globally with this expertise. Subject to our capacity limitations, we stand ready to work with both the Victorian Government and the different impacted communities in order to address the online element of vilification.

9) ANY WORK UNDERWAY TO ENGAGE WITH SOCIAL MEDIA AND TECHNOLOGY COMPANIES TO PROTECT VICTORIANS FROM VILIFICATION.

The Online Hate Prevention Institute maintains direct open channels of communications with most of the major social media companies. These relationships are multilayered with connections to their Australian staff as well as staff at the regional and global level. In some cases, we maintain a range of relationships with different staff across a company allowing us to address a range of different issues such as vilification, terrorism, foreign interference, fake news, public policy and technical innovations such as new uses of AI. In addition to direct connections to the companies we have also established a relationship with the industry group Digi who represents the companies in Australia.

Our engagement with the companies occurs in a number of different ways such as:

CRISIS RESPONSE

During and immediately following incidents such as the Bourke Street attacks (2017[26] and 2018[27]) and the Flinders Street attack (2017)[28] in Melbourne we actively monitored social media for incitement to violence. In addition to what we published openly we prepared various confidential reports in which those engaged in serious vilification and incitement to violence were identified and their activities documented. These confidential reports were shared both with police (either Victoria Police or the Australian Federal Police) and with the relevant social media

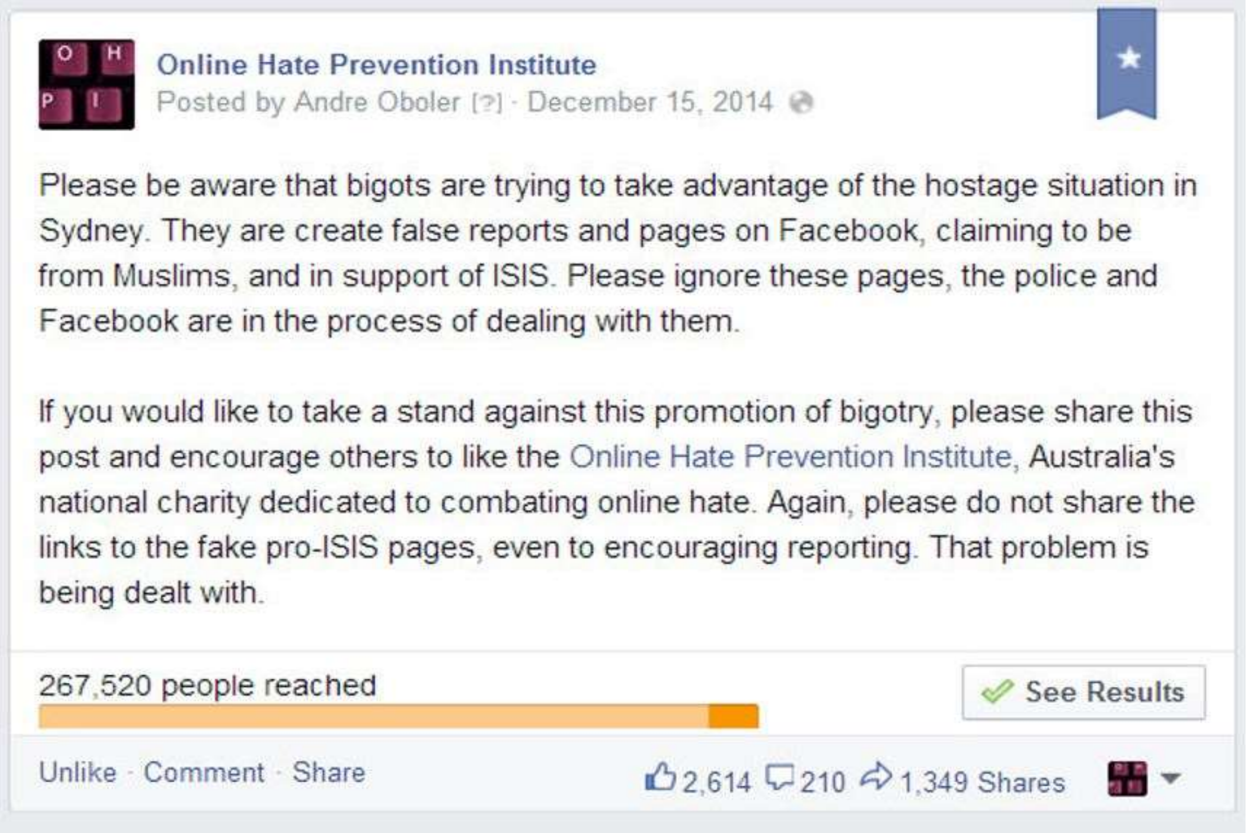
[26] A substantial report has been shared with police and is available on request. It has not be published publicly.

[27] https://ohpi.org.au/bourke-street-attack-november-2018/

[28] https://ohpi.org.au/car-attack-in-flinders-street-melbourne/

companies. We find that our requests for voluntary action on well documented situations occurring online which can put public safety at risk often receive a faster response than requests by police through formal channels.

Although physically based in Victoria, we cover all of Australia. During the Martin Place Siege in Sydney, for example, we exposed two Facebook pages setup by the far-right and pretending to be Muslim pages setup in support of the attack. The aim was to vilify Muslims and incite a race riot. We coordinated with Facebook and NSW Police to have the pages removed while also making a public announcement for people to ignore them rather than drawing attention to them as everyone who needed to be notified was now aware of them and working on it. The post was seen by over a quarter of a million people as shown in Figure 1.



Online Hate Prevention Institute

Posted by Andre Oboler [?] · December 15, 2014

Please be aware that bigots are trying to take advantage of the hostage situation in Sydney. They are create false reports and pages on Facebook, claiming to be from Muslims, and in support of ISIS. Please ignore these pages, the police and Facebook are in the process of dealing with them.

If you would like to take a stand against this promotion of bigotry, please share this post and encourage others to like the Online Hate Prevention Institute, Australia's national charity dedicated to combating online hate. Again, please do not share the links to the fake pro-ISIS pages, even to encouraging reporting. That problem is being dealt with.

267,520 people reached

Figure 1 Martin Place Public Announcement

We also find major incidents overseas can lead to serious vilification being posted by Victorians to an online audience with a significant Victorian component. Terrorist manifestos and videos of attack can also be visible to people in Victoria, which can lead to radicalisation.

We have had multiple terrorist manifestos and videos removed either globally (by the hosting company) or access blocked to Australia, including Victoria. For example, in 2019 we had two copies of manifesto from the terrorist attack in Poway (USA) removed,[29] and multiple copies of the manifesto and videos from the Halle (Germany)

[29] https://ohpi.org.au/san-diego-synagogue-attack/

terrorist attack blocked.[30] Our industry connections here cover not only social media companies but also file hosting service and website hosting companies.

MAJOR REPORTS

We have produced a number of major reports thematically examining vilification online. We try to share a draft of these reports with the companies prior to releasing them to the public. This gives them a chance to take action to address the concerns we document, and to do so with the benefit of extracted examples and our explanations of why the content is problematic.

An example of this, which has been commented on in reports by UNESCO,[31] is our report "Aboriginal Memes and Online Hate".[32] Other examples include our reports "Recognizing Hate Speech: Antisemitism On Facebook",[33] and "Islamophobia on the Internet: The growth of online hate targeting Muslims".[34]

More recently, since the launch of our advanced reporting tool "Fight Against Hate", we have been producing empirical reports measuring different kinds of online hate. Our report "Measuring the Hate: The State of Antisemitism in Social Media" gathered data on over 2,000 items of antisemitism across Facebook, YouTube and Twitter in just three months and was prepared for the Israeli Foreign Ministry.[35] Our report "Spotlight on Anti-Muslim Internet Hate Report" covered over 1000 items of anti-Muslim hate and was cited in a report of the United Nations Human Rights Council.[36] Part of this later report was supported by the Victorian Government. For these major empirical reports we have offered social media companies access to the data from their platform which we are monitoring in return for them reviewing the content a second time.

The empirical reports allow us to monitor the platforms response rate and the results have been fairly poor. We believe in the last year the effectiveness in removing content may have improved substantially, but more work is needed to verify this. We also believe this leads to those posting vilification finding new ways to work around the rules and particular around automated systems which are doing much of the removal work. Again, we are speaking to the platforms on these issues, but more research is needed. We have the tools and expertise to carry out this research and will try to raise the necessary funding to do so during 2020, however, such work requires government investment.

BRIEFINGS

We produce briefings on specific topics on a regular basis. These are usually the length of an extended blog post and they often document specific cases of vilification. We have produced almost 200 briefings since 2014.

[30] https://ohpi.org.au/hate-and-violent-extremism-from-an-online-subculture-the-yom-kippur-terrorist-attack-in-halle-germany/

[31] https://ohpi.org.au/ohpi-in-unesco-report-on-freedom-of-expression/; https://ohpi.org.au/ohpi-quoted-in-a-unesco-report-on-online-hate/

[32] http://ohpi.org.au/aboriginal-memes-and-online-hate/

[33] https://ohpi.org.au/recognizing-hate-speech-antisemitism-on-facebook/

[34] https://ohpi.org.au/islamophobia-on-the-internet-the-growth-of-online-hate-targeting-muslims/

[35] https://ohpi.org.au/measuring-antisemitism/

[36] https://ohpi.org.au/anti-muslim-hate-interim-report/

An example is a briefing on the Facebook page “Stop the mosque in Bendigo”.[37] Our briefing showcased the vilification on the page, but also provided analysis exposing the fact that only 3% of the page’s audience was from Bendigo and in total only 20% was from Victoria. Work such as this played a significant impact in correcting misperceptions of Bendigo. It also undermined a fake news narrative that sought to normalise the hate that was spread in such a group by making it appear more mainstream that it was.

Other examples include our work looking at racism and misogyny on Facebook pages about the AFL.[38] In one case we did write to our Facebook contacts leading to the worst examples on the page being removed and the page being closed soon after.

There are many more examples from our briefings, but these are best explored through our website.[39]

CONFERENCES, EVENTS AND MEETINGS

We have attended a number of conferences, symposiums and workshops along with representatives of the social media companies. We are often the only Australian civil society representative in such gatherings.

An example was the Asia Pacific launch of Tech Against Terrorism which we attended at one of the major company’s offices in Sydney. Our presentation to that high-level gathering can be seen on our Facebook page.[40] We met with a range of local and international representatives of the companies at that gathering.

We also recently met with Facebook at a conference in Singapore and previously met with Twitter and Wordpress at their headquarters in the US.

In our meetings with technology companies we often present ideas for technical changes to their platforms as well as changes to their policies. Facebook and YouTube have made significant changes to their core software as a result of our recommendations to keep people safe online. Twitter has made changes to their policy in response to discussions we have had with them.

PLANS FOR 2020

We have reached a point where we have an extensive technical capacity but limited operational budget to put it fully to work. During 2020 we are planning to run a series of month long campaigns each focused on a different area of vilification. We will be seeking sponsors for each campaign (or the series over all) as well as running crowd funding campaigns. Depending on the level of funding we can secure for each campaign, our level of activity will range from preparing a series of briefings through to producing major reports empirically monitoring the problem and the company’s response rates.

We are also running a program (funded by the Victorian Government) with the Council of Christians and Jews (Victoria). As part of this program we are training a series of community volunteers to enable them to run a two-hour community training session on tackling online vilification. The two hour training program looks at a number of

[37] https://ohpi.org.au/the-bendigo-mosque-exporting-hate-to-regional-victoria/

[38] https://ohpi.org.au/trolling-the-afl-with-racism-and-misogyny/; http://ohpi.org.au/report-adam-goodes-for-the-flog-of-the-year-page/; https://ohpi.org.au/update-on-afl-memes/

[39] https://ohpi.org.au/

[40] https://www.facebook.com/onlinehate/videos/1800284873382060/

different types of hate. We are also providing the trainers with more in-depth training into each type of hate as well as quarterly updates so they know the latest trends in online hate. We hope this will be the start of a grass roots movement empowering communities around Victoria.

We will also continue and expand our work overseas as part of the Australian Government's delegation to the International Holocaust Remembrance Alliance. As part of this a partnership with partners in Italy will see us assisting them in combating online antisemitism in Italian. We hope to connect this effort with the Victorian Italian community. We are also expanding work into Asia as a result of a partnership with a civil society organisation in Bangkok and a conference on tackling online hate in Asia which will be run mid-year. These activities increase our capacity to lead change at the global level.

We will be engaging with our contacts at the social media platforms as appropriate in relation to these activities.

APPENDIX 1: EXAMPLES OF ONLINE VILIFICATION

Figure 2 and Figure 3 shows the narrative that seeks to make vilification of Muslims acceptable by arguing it isn't racism. By making it clear vilification on the basis of religion will be treated the same as vilification on the basis of race, the Victorian legislation is an effective counter to such arguments.

Figure 4 shows an adaptation of the narrative which seeks to argue Islam is neither a race nor a religion. This is clearly much harder to argue with Islam being widely recognised as a major world religion. Victoria is ahead of the curve in sending a message that religious vilification is not acceptable.

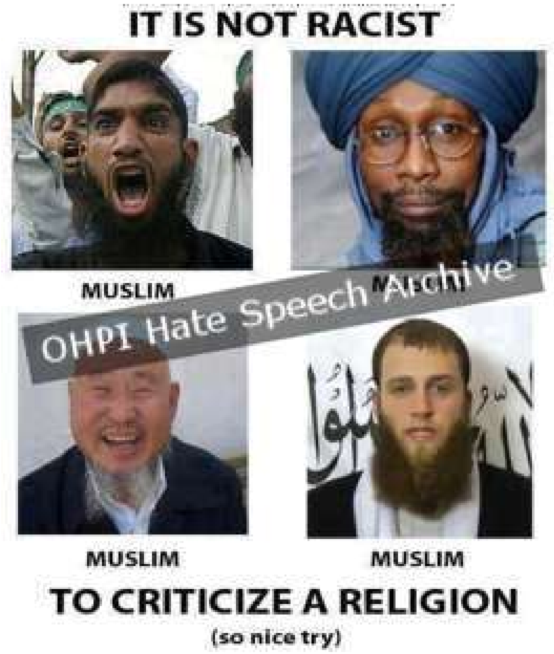
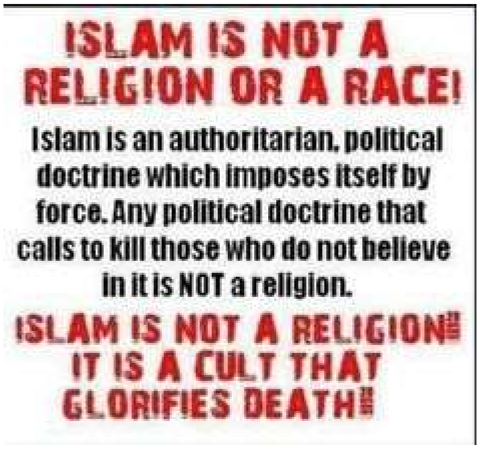


Figure 2 Meme that religious vilification of Muslims is not racism

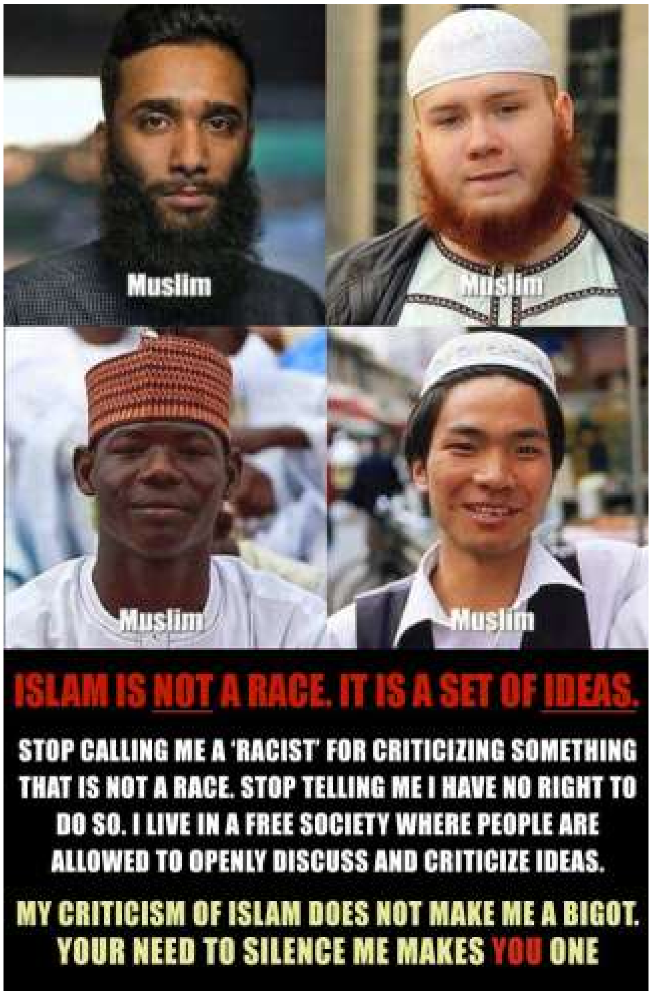


ISLAM IS NOT A RELIGION OR A RACE!

Islam is an authoritarian, political doctrine which imposes itself by force. Any political doctrine that calls to kill those who do not believe in it is NOT a religion.

ISLAM IS NOT A RELIGION!! IT IS A CULT THAT GLORIFIES DEATH!!

Figure 3 Meme promoting vilification of Muslims



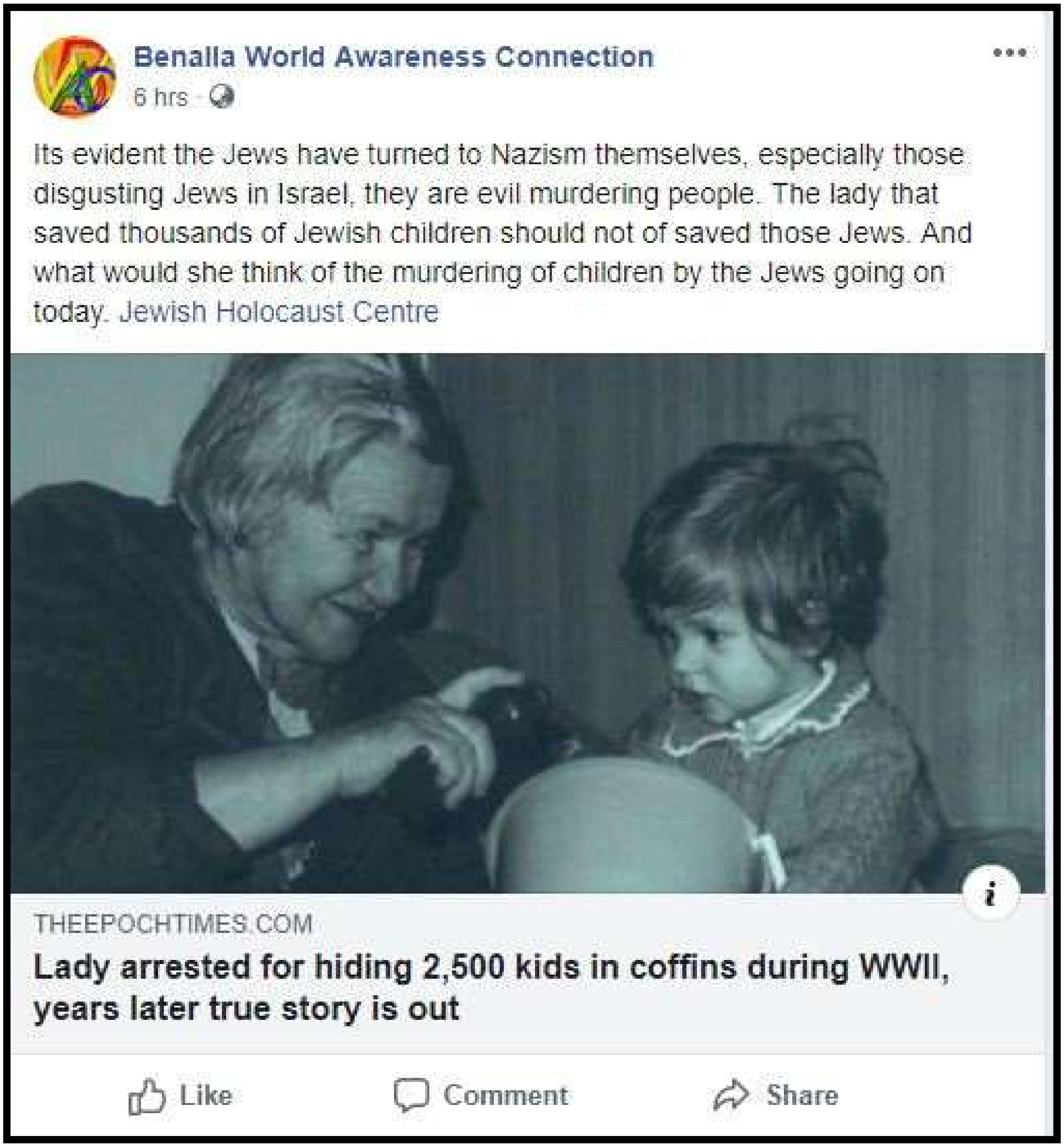
ISLAM IS NOT A RACE. IT IS A SET OF IDEAS.

STOP CALLING ME A 'RACIST' FOR CRITICIZING SOMETHING THAT IS NOT A RACE. STOP TELLING ME I HAVE NO RIGHT TO DO SO. I LIVE IN A FREE SOCIETY WHERE PEOPLE ARE ALLOWED TO OPENLY DISCUSS AND CRITICIZE IDEAS.

MY CRITICISM OF ISLAM DOES NOT MAKE ME A BIGOT. YOUR NEED TO SILENCE ME MAKES YOU ONE

Figure 4 Meme defending bigotry against Muslims

Figure 5 is an example from a Victorian based Facebook page which describes Jews as Nazis and evil and suggests Jewish children saved from the Nazis should have instead been killed.



Benalla World Awareness Connection

6 hrs ·

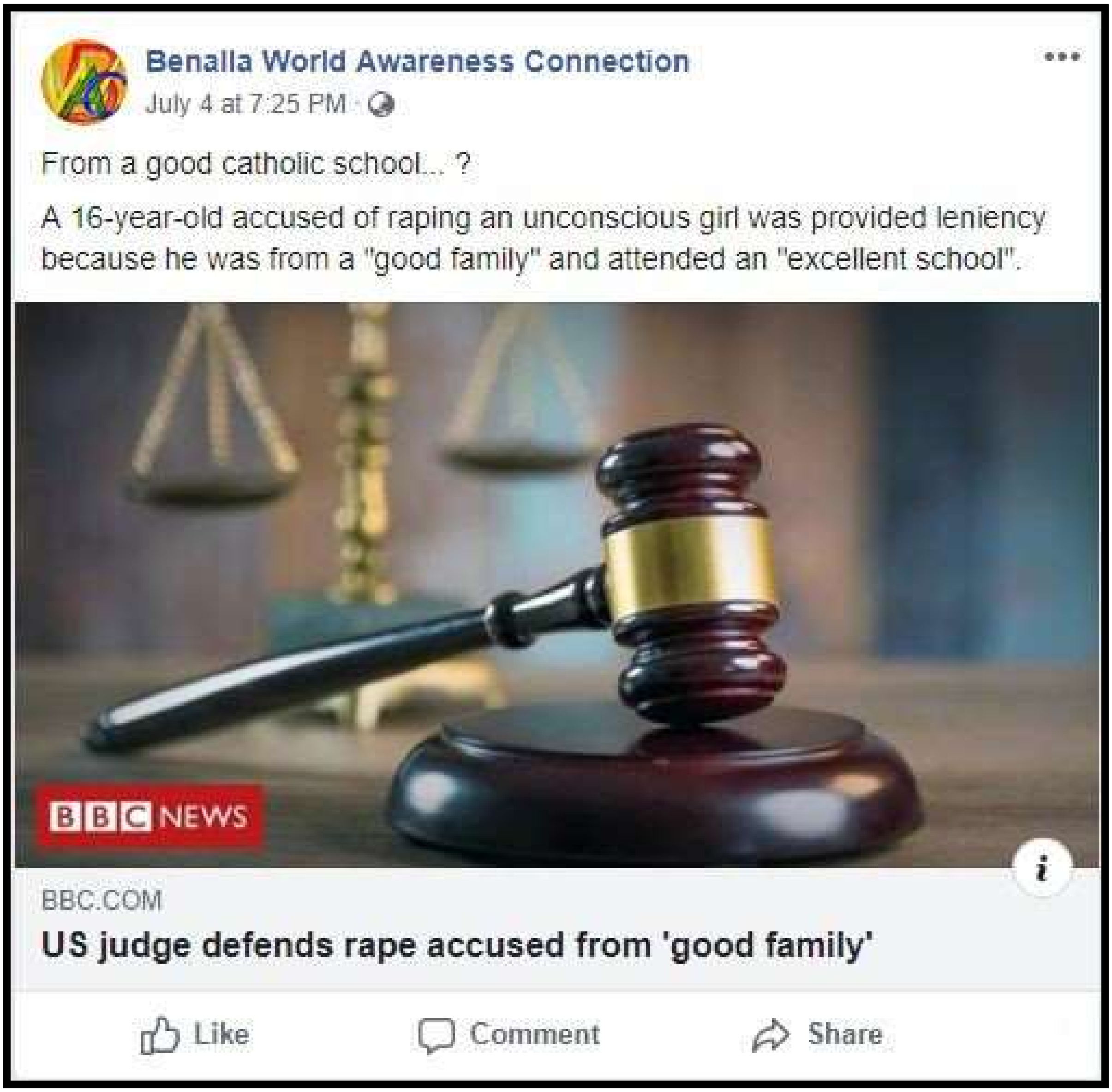
Its evident the Jews have turned to Nazism themselves, especially those disgusting Jews in Israel, they are evil murdering people. The lady that saved thousands of Jewish children should not of saved those Jews. And what would she think of the murdering of children by the Jews going on today. Jewish Holocaust Centre

THEEPOCHTIMES.COM

Lady arrested for hiding 2,500 kids in coffins during WWII, years later true story is out

Figure 5 Anti-Jewish post on Facebook

Figure 6 is an example from the same Victorian page which attacks Catholics. The article in question does not name the school or the boy and the idea the "good school" would be a Catholic school is pure conjecture designed to vilify Catholics. The page does this multiple times, usually attacking Christians in general, but posting news articles about poor behaviour and then suggesting the people are Christians – when there is no discussion of this in the underlying material.



Benalla World Awareness Connection

July 4 at 7:25 PM ·

From a good catholic school... ?

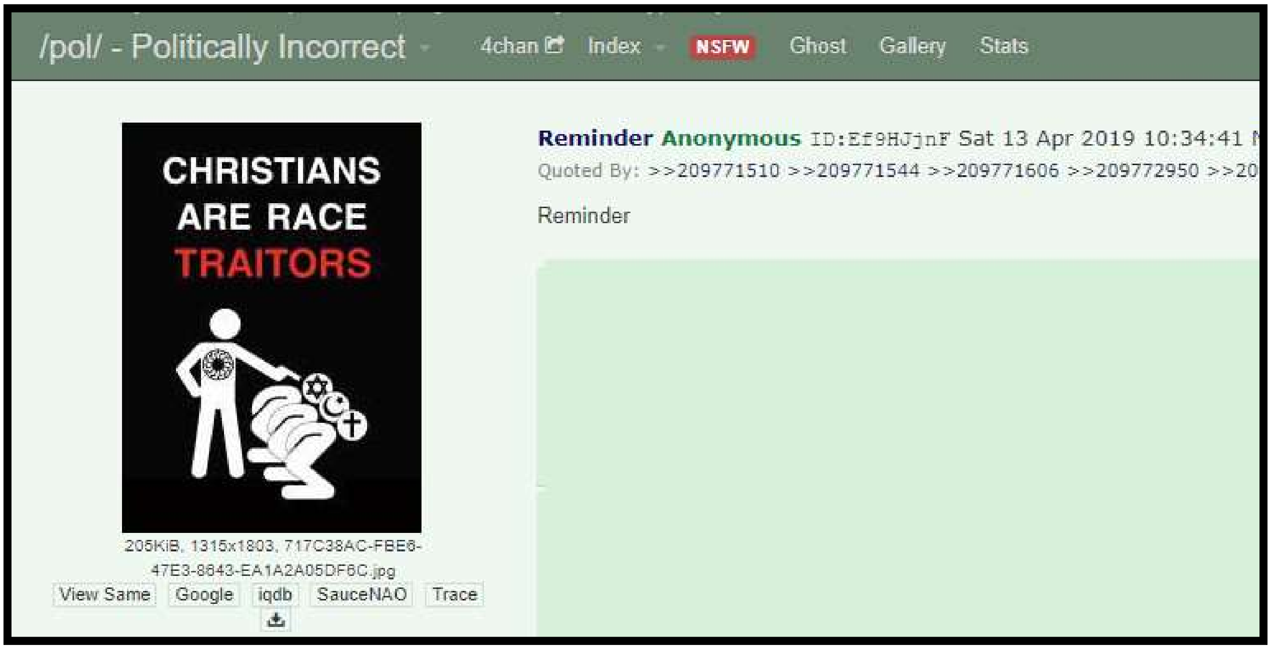
A 16-year-old accused of raping an unconscious girl was provided leniency because he was from a "good family" and attended an "excellent school".

BBC.COM

US judge defends rape accused from 'good family'

Figure 6 Anti-Catholic post on Facebook

Figure 7 puts Christians alongside Jews and Muslims as a group to be targeted with violence. This post is from /pol/ the online community behind violence like the Christchurch terrorist attack on two mosques, the attack on synagogues in Poway (USA) and Halle (Germany) and the attack focused on immigrants in El Paso (USA). More on this community which is using vilification to spreading violent extremism can be seen in our December 2019 report "Hate and Violent Extremism from an Online Subculture".[41]



205KiB, 1315x1803, 717C38AC-FBE6-47E3-8643-EA1A2A05DF6C.jpg

Reminder Anonymous ID:Ef9HJjnF Sat 13 Apr 2019 10:34:41

Quoted By: >>209771510 >>209771544 >>209771606 >>209772950 >>20

Reminder

Figure 7 Anti-Christian post on /pol/

[41] https://ohpi.org.au/hate-and-violent-extremism-from-an-online-subculture-the-yom-kippur-terrorist-attack-in-halle-germany/

APPENDIX 2: INTERNET REGULATION

The following is an extract from the report "Hate and Violent Extremism from an Online Subculture".[42]

In this report we advocate for greater cooperation between governments, technology companies and civil society. Within the context of this cooperation we believe technology platforms, with input from civil society, may choose to be proactive in removing harmful content. While the law in a particular country may not recognise a certain group as deserving of protection from hate speech, a technology platform could adopt a global position in its community standards that, nevertheless, provides such protection within the confines of that platform even in countries where there is no legal protection.

More controversially we argue that the law is the ultimate backstop and, with a few exceptions, companies that impact the people within a particular country's borders should respect the laws of that country. That is, if the law grants a particular group protection form hate speech, a technology platform should take action to prevent hate speech against that group from appearing to people in that country. We reject the idea that a platform could hold its community standards above the law when the two come into conflict. We also reject the idea that a platform could choose a jurisdiction and conform only to the laws of that jurisdiction, while ignoring the laws in the countries where its audience is based.

There are of course exceptions. There may be circumstances where a content service provider is unable to support compliance with national laws, for example, where such support would contravene laws or breach legal obligations in the content service provider's own jurisdiction. National laws may also infringe upon rights recognised in the International Covenant on Civil and Political Rights, the International Covenant on Economic, Social and Cultural Rights or similar international, regional or national instruments. Infringement in this case does not mean they strike a different permissible balance, for example between free speech and hate speech, but rather that they are entirely incompatible with such international instruments. Such exceptions should be rare, and states may well respond by seeking to block access to the platforms concerned within their territory.

It is our view that outside of exceptional circumstances, the rule of law and recognition of the sovereignty of nations requires platforms to conform to the law of the places where their audience resides. This concept is today widely accepted by major platforms, though it is implemented to varying degrees of effectiveness. The problem, which is directly relevant to this report, is how countries should handle platforms who work around their laws, for example, platforms that provide a forum for Germans to illegally glorify Nazism, or New Zealanders or Australians to access the Christchurch shooting video which both countries have declared prohibited content. Such sites may claim the continue a long history of online opposition to regulation, but they are problematic in the context of today. States may legitimately block such platforms as a last resort.

Ideally what we recommend later in this report, at Recommendation 35, is that:

> Content services should create mechanisms that enable them to restrict access to specific content on their service for users from countries where that content is illegal. This will ensure content services have the technical capacity to respect national sovereignty and comply with national laws. There may be exceptional circumstances where a content service refuses to comply with national laws, for example, if

[42] https://ohpi.org.au/hate-and-violent-extremism-from-an-online-subculture-the-yom-kippur-terrorist-attack-in-halle-germany/

the national laws conflict with customary international law, international treaties to protect human rights, or legal obligations in the content services own jurisdiction.

ORIGINS OF THE INTERNET AND ITS INHERENT RESISTANCE TO REGULATION

The early Internet grew out of ARPANET, the network of the Advanced Research Projects Agency (ARPA) established as part of the US Department of Defence to engage in blue sky research with potential military applications.[43] The system, which was essentially completed in 1972,[44] was both designed and used by high-profile researchers working in elite institutions.[45] There was a collaborative ethos by those building the system and a strong demands for modifications and innovation which went beyond the intended purpose of the system.[46] Responsibility for the network was transferred to the Defense Communications Agency in 1975, and in 1983 the agency split the system creating MILNET for military communications while ARPANET would continue to support research.[47]

The shift in control back to universities and research institutions and was “an essential first step towards achieving ARPA’s goal of transferring the network to civilian control”.[48] Access to ARPANET was limited to certain institutions and which was seen as “increasingly perceived as irritating and dysfunctional” by those in the developing field of Computer Science.[49] This led to the creation of the Computer Science Network (CSNET) in the early 1980s by US National Science Foundation (NSF) and its infrastructure eventually became the backbone to ARPANET as well.[50] Commercial use of the network was prohibited under National Science Foundation’s ‘acceptable use’ policy.[51]

By 1994 the National Science Foundation decided the network needed to be privatised in order to allow commercial exploitation.[52] In privatising the network, the decision was made to have many smaller companies, Internet Service Providers, cooperate in running the backbone rather than entrusting the system as a whole to a major technology or telecommunications company.[53] The emerging technology and culture developed in a manner that deliberately sought to resist centralised control and government control in particular.

The Internet had finally, after significant effort, broken away from the restrictive control that had been applied by various parts of government. The early commercial Internet was “essentially a geek preserve, with a social ethos that was communal, libertarian, collaborative, occasionally raucous, anti-establishment and rich in debate and

[43] John Naughton, 2016. “The evolution of the Internet: from military experiment to General Purpose Technology”, *Journal of Cyber Policy*, volume 1, number 1, pp. 5-28, p. 7.

[44] Ibid 8.

[45] Ibid.

[46] Ibid 9.

[47] Ibid 10.

[48] Ibid 11.

[49] Ibid 11.

[50] Ibid.

[51] Ibid.

[52] Ibid 12.

[53] Ibid.

discussion".[54] The ethos of the early Internet is best displayed in "A Declaration of the Independence of Cyberspace" posted by John Perry Barlow on February 8th, 1996:

> "Governments of the Industrial World, you weary giants of flesh and steel, I come from Cyberspace, the new home of Mind. On behalf of the future, I ask you of the past to leave us alone. You are not welcome among us. You have no sovereignty where we gather… You do not know our culture, our ethics, or the unwritten codes that already provide our society more order than could be obtained by any of your impositions."[55]

This 1990s view sees the Internet as having an "exceptional" nature which made it not susceptible to regulation by the laws of nation states.[56] Some legal scholars went as far as to argue that the Internet has its own sovereignty and should have its own laws that reflected its "special character".[57] The view that the internet was something apart and needed protection from government regulation also gained ground in the courts. In *ACLU v Reno* (1996) US Federal Judge Stewart Dalzell wrote that, "[a]s the most participatory form of mass speech yet developed, the Internet deserves the highest protection from government intrusion. …The absence of governmental regulation of Internet content has unquestionably produced a kind of chaos, but as one of plaintiffs' experts put it with such resonance at the hearing: 'What achieved success was the very chaos that the Internet is. The strength of the Internet is that chaos'".[58]

THE DEATH OF INTERNET EXCEPTIONALISM

Jack Goldsmith and Tim Wu, writing in 2006, considered the 1990s perception of Internet exceptionalism in which "many believed that nations could not control the local effects of unwanted Internet communications that originated outside their borders, and thus could not enforce national laws related to speech, crime, copyright, and much more."[59] Reflecting on the decade since 1996 which had "shown that national governments have an array of techniques for controlling offshore Internet communications, and thus enforcing their laws, by exercising coercion within their borders," they rejected this view and warned of the death of Internet exceptionalism.[60]

[54] Ibid 12.

[55] John Perry Barlow, "A Declaration of the Independence of Cyberspace," at https://www.eff.org/cyberspace-independence, accessed 18 December 2019.

[56] Jack Goldsmith and Tim Wu, 2006. *Who Controls the Internet: Illusions of a Borderless World.* New York: Oxford University Press, p. viii.

[57] David Post and David Johnson, 1996. "Law and Borders – The Rise of Law in Cyberspace," *Stanford Law Review*, volume 48, pp. 1367-1402, Online at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=535, accessed 18 December 2019, pp. 1400-1401.

[58] *American Civil Liberties Union v. Reno*, 929 F. Supp. 824, 883 (ED Pa. 1996), at https://law.justia.com/cases/federal/district-courts/FSupp/929/824/1812782/.

[59] Jack Goldsmith and Tim Wu, 2006. *Who Controls the Internet: Illusions of a Borderless World.* New York: Oxford University Press, p. viii.

[60] Ibid.

Goldsmith and Wu predicted the internet would become bordered, splitting apart to conform to local conditions, including language, content and norms.[61] They predicted the Internet would "differ among nations and regions that are increasingly separated by walls of bandwidth, language and filters", reflecting "top-down pressures from governments that are imposing national laws on the Internet within their borders" and "bottom-up pressures from individuals in different places who demand an Internet that corresponds to local preferences", as well as by the efforts of technology companies to meet those demands.[62] While noting that many will "lament the death of the borderless Internet", Goldsmith and Wu state that "the geographically bordered Internet has many underappreciated virtues", including meeting the demands of the citizenry that governments prevent them from harming each another, and to protect them from harm from abroad.[63] They argue that the "bordered Internet accommodates real and important differences among people in different places, and makes the Internet a more effective and useful communications tool as a result".[64]

Goldsmith and Wu note that as "governments increase their control, they replicate their vices on the Internet".[65] They discuss China's effort at political control and economic self-aggrandisement, but also the risk in democratic countries of "corruption and imperfections of the political process".[66] These potential problems did not dissuade them from the view that "on the whole decentralized rule by nation-states reflects what most people want."[67]

The shift to a "bordered Internet" was significantly slowed with the rise of Web 2.0 and social media. Those responsible for placing harmful content on the Internet no longer needed their own domain or physical servers. Their IP address and location would often be masked by the technology companies. The argument for shutting down a website when the owner was using it for harmful activities after refusing to desist or take remedial action was greatly weakened when it was not the site owner but instead the visitors to the site who engaged in harmful activities or uploaded harmful content. The idea of penalising the company and other users of the platform for actions of a small minority abusing the technology, for example, by taking down a service for non-compliance, was seen as a disproportionate response. Technology companies also argued they were incapable of taking effective action given the volume of content on their services, or that it would be prohibitively expensive, and that efforts to increase obligations on them would therefore stifle innovation.

Technology companies also sought to maintain a unified approach across their platforms. The bottom up pressure Goldsmith and Wu expected, where individuals would want the Internet to reflect their local preferences and companies would seek to meet this demand, was strongly resisted by the growing companies. This was most evident in Facebook's resistance to banning Holocaust denial in spite of public calls for such measures, and even in countries where such content was illegal.[68] Their initial position was that country specific rules were not possible,

[61] Ibid.

[62] Ibid.

[63] Ibid.

[64] Ibid.

[65] Ibid.

[66] Ibid ix.

[67] Ibid.

[68] Andre Oboler, 2009. "Facebook, Holocaust Denial, and Anti-Semitism 2.0", *JCPA: Post-Holocaust and Anti-Semitism*, August 2009. Online at http://jcpa.org/article/facebook-holocaust-denial-and-anti-semitism-2-0/, accessed 12 December 2019.

though they later revised this position saying they would block access to content in countries where the content was illegal. As their spokesperson, Barry Schnitt, explained:[69]

> "When dealing with user generated content on global websites, there are occasions where content that is illegal in one country, is not (or may even be protected) in another. For example, homosexual content is illegal in some countries, but that does not mean it should be removed from Facebook. Most companies approach this issue by preventing certain content from being shown to users in the countries where it is illegal and that is our approach as well. We have recently begun to block content by IP [the "address" of a computer on the internet] in countries where that content is illegal, including Nazi-related and holocaust denial content in certain European countries."

This set the precedent,[70] and the same approach was used in blocking access to the 'Everybody Draw Mohammed Day!' Facebook page in Pakistan,[71] and in India.[72] The approach was, however, inconsistent in its application. While German hate speech laws saw refugees as a protected group, Facebook's community standards, as they stood in 2015, did not.[73] The gap led to anti-immigrant content remaining online for weeks, or never being removed, much to the frustration of the German Government.[74] The online hate was linked by researchers to a rise in offline violence against refugees.[75] In February 2016 Facebook admitted it had made a mistake, Mark Zuckerberg

[69] Ibid.

[70] Helen A.S. Popkin, 2010. "Free Speech vs. Hate Speech on Facebook," *NBC Connecticut* (19 May), at https://www.nbcconnecticut.com/news/politics/free_speech_vs__hate_speech_on_facebook/1864429/, accessed 19 December 2019.

[71] Reuters, 2010. "Facebook admits censoring content in Pakistan," *Reuters* (2 June), at https://www.reuters.com/article/urnidgns852573c400693880002577350003099ea-idUS126971660620100601, accessed 19 December 2019.

[72] John Ribeiro, 2010. "Facebook prevent Indian access to anti-muslim group," *IDG News Service* (24 May), at https://www.cio.co.uk/it-leadership/facebook-prevent-indian-access-to-anti-muslim-group-3224505/, accessed 19 December 2019.

[73] Amar Toor, 2015. "Facebook will work with Germany to combat anti-refugee hate speech", *The Verge* (15 September), at https://www.theverge.com/2015/9/15/9329119/facebook-germany-hate-speech-xenophobia-migrant-refugee, accessed 19 December 2019.

[74] Katrin Bennhold, 2018. "Germany Acts to Tame Facebook, Learning From Its Own History of Hate", *The New York Times* (19 May), at https://www.nytimes.com/2018/05/19/technology/facebook-deletion-center-germany.html, accessed 22 August 2019.

[75] Karsten Müller and Carlo Schwarz, 2018. "Fanning the Flames of Hate: Social Media and Hate Crime", Working Paper Series: Centre for Competitive Advantage in the Global Economy, Warwick University. Online at https://warwick.ac.uk/fac/soc/economics/research/centres/cage/manage/publications/373-2018_schwarz.pdf, accessed 19 December 2019.

apologised saying that, "learning more about German culture and German law has led us to change our approach" and refugees became a protected group on Facebook.[76]

ASSERTIONS OF SOVEREIGNTY AND THE SHIFT TO GOVERNMENT REGULATION

Following the introduction of a voluntary agreement between major technology platforms and the German government, tests were carried out by the government to assess the level of compliance in removing hate speech reported by regular users.[77] The results were disappointing, with one test showing a 46% removal rate, and the other just 39%.[78] The German government then introduced the *Network Enforcement Law* which outlined 21 types of "manifestly illegal" content which platforms were required to quickly remove.[79] The law, passed in June 2017, requires platforms to remove manifestly illegal content within 24 hours if its illegality is obvious, or within 7 days if a determination on the nature of the content is more difficult.[80] The law provides for fines of up to €50 million for non-compliance.[81]

Facebook said "It is perfectly appropriate for the German government to set standards", but argued that it, Facebook, did *not* want to be the arbiter of what breached the standards.[82] German officials rejected this by arguing that the platforms were already the arbiters when it came to compliance on their platform.[83] Gerd Billen, the secretary of state for Germany's Ministry of Justice and Consumer Protection, said that the question was "Who is sovereign? Parliament or Facebook?"[84] This highlights that this was not a negotiation on how to proceed (as occurred when voluntary agreements were created), but an assertion of the rights and powers of state sovereignty.

A similar assertion of sovereignty was made by the Australian Government in 2019 following the Christchurch attack when new criminal provisions were created with significant penalties for technology platforms, whether inside or outside Australia, that failed to expeditiously remove 'Abhorrent Violent Material' they made visible in Australia. 'Abhorrent Violent Material' is a term defined in the legislation which included video recorded by a terrorist of their violent attack. Australia has also been active in asserting its rights over taxation, passing the *Multinational Anti-Avoidance Law* in 2015 to "ensure that multinationals pay their fair share of tax on the profits

[76] Christina Beck, 2016. "Mark Zuckerberg confronts 'hate speech' in Germany and at Facebook", *Christian Science Monitor* (27 February), at https://www.csmonitor.com/USA/Society/2016/0227/Mark-Zuckerberg-confronts-hate-speech-in-Germany-and-at-Facebook, accessed 19 December 2019.

[77] Katrin Bennhold, 2018. "Germany Acts to Tame Facebook, Learning From Its Own History of Hate", *The New York Times* (19 May), at https://www.nytimes.com/2018/05/19/technology/facebook-deletion-center-germany.html, accessed 22 August 2019.

[78] Ibid.

[79] Ibid.

[80] AFP, 2017. "Germany imposes €50 million fines on social media firms that don't delete hate speech", *The Local.de* (30 June), at https://www.thelocal.de/20170630/germany-imposes, accessed 12 December 2019.

[81] Ibid.

[82] Katrin Bennhold, 2018. "Germany Acts to Tame Facebook, Learning From Its Own History of Hate", *The New York Times* (19 May), at https://www.nytimes.com/2018/05/19/technology/facebook-deletion-center-germany.html, accessed 22 August 2019.

[83] Ibid.

[84] Ibid.

earned in Australia".[85] On introducing the law the government explained that, "some multinational entities engage in deliberate tax avoidance, exploiting legal loopholes to pay less tax than the law intended".[86] Google is the latest to reach a settlement with the Australian Taxation Office after agreeing in December 2019 to pay a $481.5 million settlement, this follows previous settlements by Apple, Facebook and Microsoft. [87]

Sir Tim Berners Lee, the inventor of the World Wide Web, has also supported the notion of government intervention. In 2018 he called for a "legal or regulatory framework that accounts for social objectives".[88] He warned that the companies society was relying on to fix a growing list of online problems were "built to maximise profit more than to maximise social good",[89] and that the Web itself has changed and was now "compressed under the powerful weight of a few dominant platforms".[90]

In March 2019, Facebook's founder, Mark Zuckerberg, took a similar position writing, "I believe we need a more active role for governments and regulators. By updating the rules for the Internet, we can preserve what's best about it — the freedom for people to express themselves and for entrepreneurs to build new things — while also protecting society from broader harms."[91] In the area of hate speech, however, he went on to call for a more standardised approach and for "third-party bodies to set standards governing the distribution of harmful content and to measure companies against those standards" to ensure the volume of hate that remained online was minimized.[92] While we support the approach we note the lack of localisation to national laws. By contrast, he was very direct in saying "legislation is important for protecting elections".[93] It highlights that at least for hate speech, there is still a push for global rules, but with some greater engagement by governments, even as some exceptions based on national law emerge in countries like Germany and now potentially France.

Ultimately governments have the power and authority to regulate online activities that have an impact within their borders. Their power comes from their ability to make and enforce laws, and their authority results from their

[85] Australian Taxation Office, 2017. "Combating multinational tax avoidance – a targeted anti-avoidance law," *Australian Taxation Office* (10 August), at https://www.ato.gov.au/Business/International-tax-for-business/In-detail/Doing-business-in-Australia/Combating-multinational-tax-avoidance---a-targeted-anti-avoidance-law/, accessed 23 December 2019.

[86] Explanatory Memorandum to the *Tax Laws Amendment (Combating Multinational Tax Avoidance) Bill* 2015, at https://parlinfo.aph.gov.au/parlInfo/download/legislation/ems/r5549_ems_f2f9c061-45d9-4f1f-b8f9-eb140cdc08ae/upload_pdf/503830.pdf;fileType=application/pdf, accessed 23 December 2019.

[87] Jack Gramenz, 2019. "Google pays $481.5 million settlement to tax office but still doesn't think it's done anything wrong," *News.com.au* (19 December), at https://www.news.com.au/technology/online/google-pays-4815-million-settlement-to-tax-office-but-still-doesnt-think-its-done-anything-wrong/news-story/f4468d258c007637524bc5ee27989cf4, accessed 23 December 2019.

[88] Tim Berners Lee, 2018. "The web is under threat. Join us and fight for it," *Web Foundation* (12 March), at https://webfoundation.org/2018/03/web-birthday-29/, accessed 19 December 2019.

[89] Ibid.

[90] Ibid.

[91] Mark Zuckerberg, 2019. "Mark Zuckerberg: The Internet needs new rules. Let's start in these four areas." *The Washington Post* (31 March), at https://www.washingtonpost.com/opinions/mark-zuckerberg-the-internet-needs-new-rules-lets-start-in-these-four-areas/2019/03/29/9e6f0504-521a-11e9-a3f7-78b7525a8d5f_story.html, seen 23 December 2019.

[92] Ibid.

[93] Ibid.

sovereignty. In a digitally connected world, it is increasingly evident that the idea of a country's sovereign territory is being reinterpreted to include online communications with people within the country's physical territory. This is the only way nations can meet the increasing demands of their citizens for an online experience which takes account of the "real and important differences among people in different places" which Goldsmith and Wu highlighted.[94]

DOCTRINAL PRINCIPLES OF INTERNET REGULATION

As governments move to regulation there are a number of legal doctrinal principles which ought to be considered.[95] Some of the key concepts are outlined here and are considered in the recommendations made in this report.

The *principle of generality* which holds that it is better to have laws that apply in all circumstances rather than laws that are specific to the online context.[96] Under this principle it would be better, for example, to prohibit the dissemination of a terrorist manifesto than to prohibit the hosting of a manifesto in an online service.

The *principle of inclusion* suggests that it should not be possible to escape the law by acting online rather than in the real world.[97] This creates a need for laws that enable technical solutions to overcome what would otherwise be technical barriers in applying the general law. Such laws might, for example, make it easier to identify an online user or require the preservation of digital evidence.

The *principle of appropriate adaptation* states that laws targeted at the Internet specifically are appropriate when there is an "impact on the nature of the conduct or its prevalence" as a result of harmful behaviour going online.[98] The risk of content inciting violence spreading online to a large audience, creating a significant likelihood that it would be seen by someone susceptible to the message and willing to act, under this principle would justify special laws to ensure rapid removal of such content in order to contain the spread and reduce the risk.

As a corollary to the principle of appropriate adaptation, the nature and prevalence of conduct online can at times make non-criminal responses in practice more effective than a criminal response.[99] Where sanctions by online platforms can discourage negative behaviour, for example the posting of hate speech, these sanctions ought to be preferred to legal remedies. The volume of problems that need to be addressed might otherwise overwhelm the justice system. It is this principle which suggests platforms should make the initial call on classifying content as the volume of decisions their business model creates exceeds what the justice system can handle.

[94] Jack Goldsmith and Tim Wu, 2006. *Who Controls the Internet: Illusions of a Borderless World.* New York: Oxford University Press, p. viii.

[95] Andre Oboler, 2014. "Legal Doctrines Applied to Online Hate Speech", at http://www.austlii.edu.au/au/journals/ANZCompuLawJl/2014/4.pdf, accessed 12 December 2019.

[96] Jonathan Clough, 2010. *Principles of Cybercrime*. Cambridge: Cambridge University Press, p. 15.

[97] Neal Kumar Katyal, 2001. "Criminal law in cyberspace", *University of Pennsylvania Law Review,* volume 149, pp. 1005-1007, p. 1003.

[98] Jonathan Clough, 2010. *Principles of Cybercrime.* Cambridge: Cambridge University Press, p. 16.

[99] Andre Oboler, 2014. "Legal Doctrines Applied to Online Hate Speech", at http://www.austlii.edu.au/au/journals/ANZCompuLawJl/2014/4.pdf, accessed 12 December 2019.

The *principle of necessary criminalisation* holds that when responses short of the criminal law would be 'ineffective, impractical or insufficient', a criminal response is justified.[100] The conclusion then is that "the criminal law is needed as a final response to online hate speech",[101] and represents the endgame of a linear series of escalating responses.

[100] Jonathan Clough, 2010. *Principles of Cybercrime.* Cambridge: Cambridge University Press, p. 16.

[101] Andre Oboler, 2014. "Legal Doctrines Applied to Online Hate Speech", at http://www.austlii.edu.au/au/journals/ANZCompuLawJl/2014/4.pdf, accessed 12 December 2019.