AUSTRALIAN HATE CRIME NETWORK

SUBMISSION TO:

Australian Government, Department of Infrastructure, Regional Development and Communications

on the Online Safety Bill's Exposure Draft.

BY: Australian Hate Crime Network ('the Network')

DATE: 12 February 2021

DISCLAIMER: The contents of this submission represent the views of non-government and academic members of the Australian Hate Crime Network. It does not represent the views of representatives of any federal or state government agency or department.

ABOUT: the Australian Hate Crime Network

The AHCN is a partnership composed of three sectors of society: academics, representatives of minority communities, and as associates, people from relevant government departments, including police. AHCN recognises that hate crime may be committed on the grounds of, but is not limited to, race, religion, ethnic or national origin, sexuality, gender, gender identity, age, disability or homelessness status.

The AHCN aims to:

1. provide leadership, advocacy and support for state and national government responses to hate crime and hate incidents;

2. provide an educative and advisory role to key agencies and services on preventing and responding to hate crime and hate incidents;

3. enhance community awareness of hate crime and hate incidents, and encourage reporting, help seeking and access to available resources;

4. monitor and review patterns in hate crime and hate incidents;

5. advocate for improvement in data collection, law enforcement and criminal justice responses; and,

6. collect and distribute relevant current research and knowledge on hate crime and hate incidents.

AHCN webpage:
https://www.sydney.edu.au/law/our-research/research-centres-and-institutes/australian-hate-crime-network.html

AHCN Facebook page:
https://www.facebook.com/AustralianHateCrimeNetwork

OPENING REMARKS

The Australian Hate Crime Network (AHCN) welcomes the opportunity to comment on the Online Safety Bill's Exposure Draft.

On the whole, the Bill represents a formidable step forward in establishing oversight systems for digital platforms.

We welcome the inclusion of a cyber-abuse complaints scheme for adults, alongside the cyber-bullying scheme for children, as well as the 24 hour timeframes imposed on platforms to take down material. However, we recommend that social media sites are made, via this Act, to delete posts and comments that vilify or incite violence against people on the grounds of their race, religion, gender, gender identity, sexuality or disability, and delete the accounts of users who repeatedly violate those rules.

We welcome the intimate image abuse scheme, as well as the cultural and religious considerations that it includes.

We welcome the expansion of the role of the e-Safety Commissioner, including its powers to request Industry Codes and set Industry standards under the Online Content Scheme, along with its investigative and decision-making powers in relation to complaints. We welcome the civil penalties for platforms. However, we query the extent to which public harms such as vilification, and specifically dehumanisation of 'outgroups' in the context of extremism, incitement to violence or glorification of genocide, and disinformation, will then fall within the remit of such Codes and practices as they do not clearly form part of the Content Scheme.

Likewise, we support the integration of the Basic Online Safety Expectations ("the Expectations") with data reporting requirements for platforms, as a way to increase accountability in priority policy areas. However, we have some recommendations about how vilification and incitement to violence should be included in the Expectations. It is possible within these powers for the e-Safety Commissioner to request data for their Office in relation to breaches of platforms' terms of service, which could include hate speech and vilification. This does however rely on the e-Safety Commissioner taking an interest in this area. Data over the longer term is necessary to evaluate trends, and therefore it would make sense to include vilification and incitement to violence explicitly in the Expectations from the beginning. It would also make it clearer to Australians what is unacceptable and what can be reported.

We welcome the power to request the contact details connected to an online account as part of the e-Safety Commissioner's conferred investigative powers, along with the Commissioner's power to impose penalties on platforms for non-compliance. This is crucial to removing barriers to the enforcement of justice online. It is particularly useful where members of the Australian community are targeted by abuse and threats. However, it is less clear whether this power could be invoked where an aggrieved party is trying to bring a complaint against an online actor, where the actor is allegedly vilifying an identifiable group protected at law. As mentioned earlier, vilification as a civil action does not appear to fall within Content Scheme, and therefore it is less clear to what extent the Commissioner could

assist in the administration of civil justice, even where it connects clearly to the safe experience of online users.

We welcome the move to make it possible for the Commissioner to request that material by terrorists, such as Brenton Tarrant's manifesto, be taken down from a platform and to issue penalties for non-compliance. It is also noted that under this new framework, the Commissioner is not limited to taking action on content that has been formally RC Classified. To enable fast action, the Commissioner may act where it deems the material to be equivalent to RC classified material.

The Australian Hate Crime Network is also aware of concerns about the potential for this particular power to be weaponised against LGBTIQ communities, who have previously expressed concerns about the classification of consensual sex acts as RC Classified material under the Classification (Films, Publications and Computer Games) Act. It may be helpful to qualify that this power cannot be used in a manner that is discriminatory in its effect to one particular community on the basis of a protected characteristic; or to interfere with health promotion or education online in regard to acts that are RC Classified material.

We welcome the takedown, blocking and link deletion request powers and penalties for non-compliance in relation to abhorrent violent material connected to crisis events, especially given its potential to dehumanise, normalise and induce further violence based on discriminatory hatred. The Australian Government's swift action on this particular type of harm following the Christchurch terror attack's livestreaming on Facebook is to be commended.

However, there are serious ongoing harms not contemplated by the Act, addressed below.

FEEDBACK

From a community perspective, there is a strong need to see more steps taken upstream, to prevent violence from occurring. Australia has extensive counterterrorism laws to identify those who plan or execute terrorist attacks, and those who advocate in support of terrorism, but this policy approach ignores that extremism is part of the continuum of hate crimes. With respect to 'lone actor' terrorism, hate crimes can be a useful early threat indicator. Individuals who adopt a violent or racist ideology, generally express their beliefs through words and low-level actions that do not amount legally to extremist actions or terrorism but do meet the standard of a hate crime or hate incident. In threat management the concept of leakage is important as it indicates that a person of concern is telegraphing his or her intentions. Hate crimes act in a similar way, telegraphing that an individual is on the road to extremism or identifying those on the path of escalation to violent action.

Through the development of a national hate crime response, individuals and groups can be identified in the early stages and effective intervention measures can be employed to greatly reduce any risk they may pose. In the online context, that hate crime response must include repercussions for serially posting vilifying or inciteful content in echo chamber environments, which, for reasons described later, is best implemented through the administrative powers of the Online Safety Act.

Extremist movements have an inordinate role in fostering disharmony. In hate crimes roughly 90%-95% of all hate crime offenders have no relationship to an organised hate group (or extremist group), but, despite this, organised hate groups play an important role in driving behaviour. The mere existence of a group espousing an ideology gives credence to the feelings and beliefs of an individual. In addition, groups offer psychological and moral support, if not material support, as they reaffirm that the individual is not alone in their beliefs. Groups can offer moral support directly, even if the individual does not join the group. Having a point of contact, especially in the virtual world, can sustain and develop an individuals' beliefs and potential for violent action.

Groups and individuals may not meet the benchmark to be proscribed as a terrorist organisation or terrorist, but they may invite individuals down that path. Since about 2017, right wing groups have moved away from reliance on a group dynamic, advocating and promoting 'lone actor' activism instead. This has been heavily promoted through the works of the likes of James Mason and his collected works, 'Siege'. Groups will always exist but dependence on them is no longer critical. The threat comes from loosely affiliated individuals. The INCEL (INvoluntary CELibate) movement is an example of this. Whilst INCEL is a movement, it is loosely based with no set leadership structure or command and control that typically characterises formalised groups.

In the 21st century this influence has become easier with the use of technology and the ability to create virtual 'echo chambers' around issues. This capability has removed the limitation of physical groups and especially has reduced the impediment of distance. Individuals in Australia are capable of being influenced 24 hours a day by like-minded individuals in the United States, Europe, the United Kingdom, and the Middle East.

Whilst the aim of disruption of hate speech and the use of symbols is a noble aim, it must be tempered with reality. At times, banning has the opposite effect to that which is intended, making the banned object a taboo and therefore more desirable than it was before. An effective response must be a balanced response, for if it is not balanced, we can play into the hands of the extremists.

A popular approach to not only address the harmful impact of hate speech and the use of symbols is to focus on incitement to violence. Symbols have a powerful impact and are often used to incite action, whilst speech when used effectively can have a powerful effect not only on the psyche but also bringing action to life. The saying that words have power is true and recent events in the United States bear testament to this. Symbols or speech that incite violence or dehumanise an 'out' group is an approach that, when done properly, has been effective in curbing the ability of extremists to cause disruption. Even in the United States where the First Amendment protects speech, incitement is addressed through the recognition that 'fighting words' are not protected. Incitement is a difficult term in the current Australian legal environment as the definition is not clear and the legal benchmark moves from case to case. The lack of a fixed and useable definition has prevented the effective use of legislation that requires proof of incitement. We are cognisant that this may present some barriers to the e-Safety Commissioner, who acting in an administrative capacity would need to adjudicate on more 'clear cut' cases. That issue is even more alive when referring to vilification, which carries a number of free speech defences and requires even more contextual consideration.

An example of effective incitement legislation can be found in the Canadian Criminal Code under section 318 to 320. This legislation makes it a criminal offence to advocate or promote genocide (Section 318), communicate statements in a public place inciting hatred against any identifiable group where such incitement is likely to lead to a breach of the peace (Section 319(1)) and communicate, except in private conversation, statements that wilfully promote hatred against an identifiable group (Section 319 (2)). Finally, Section 320 and 320.1 allows judges to issue an order to confiscate hate propaganda in any form, including data on a computer system. This legislation clearly defines the terms and the Court rulings have further clarified these terms, allowing for a clear direction on the use of the legislation. The advantage of adopting the Canadian approach is that the Canadian legal system is based on the same legal system as Australia, making it easily transferable.

One of the hardest concepts to navigate is that of free speech. In a democracy it is important that individuals and groups have the right to speak freely, even if those views are unpopular and in the minority. There needs to be a balanced approach that ensures differing views can be expressed and that no member of a society is isolated and targeted for violence for being who they are. Striking this balance is a difficult act but must be achieved.

One factor that must be considered is what is deemed to be a hate symbol. Many symbols that are identified with organised hate groups have been co-opted from legitimate religious and cultural groups. Many of the symbols synonymous with white supremacy have legitimate uses, for example, the Celtic Cross (known as the White Power cross) is a pagan symbol used by those who practice and follow the Celtic traditions. The Southern Cross is used by right-wing groups in Australia but does not mean that a person with the symbol is associated with

or subscribes to extreme right-wing groups and their ideology. The Southern Cross forms part of the Eureka Flag, a symbol that is used by both left and right-wing groups and individuals.

It should be noted that individuals who have a legitimate use for symbols may also hold extremist views, further muddying the waters regarding the purpose of using a symbol, thereby making prosecution for offences complex. Individuals who hold pagan beliefs, like Asatru, use many of the symbols of the extreme right for non-extremist purposes. However, this religion has been adopted by many in the white supremacist community, leading to a legal defence for the use of symbols. The key is context, and if a legislative redress is the chosen course, consideration must be given to the potential difficulties in prosecuting any offences. As outlined previously there is an inherent danger if the first response is to ban. Not only does it risk capturing individuals who have no malice in the use of the symbols (with the potential to undermine the effectiveness of the legislation) but also risks driving the very behaviour it was intended to prevent.

No matter what action is taken there are fundamentals that must be addressed. The most important of these is that any legislative steps must be applied equally and without prejudice to all individuals, groups, and ideologies. The most dangerous thing is if the beliefs of one group are seen to be targeted but the beliefs of another group or ideology, that are just as dangerous, are treated as acceptable or excuses are made for the ideology. All extremist ideologies pose a threat, just not the one that is the political and media favourite.

The display of 'white power' advocacy, and a cross-burning, in the Grampians National Park in Victoria in late January 2021 by a neo-Nazi group, National Socialist Network (NSN), as reported in the mainstream media, provides an example of extremist activity. That activity falls into a gap as technically the group was not urging violence against a minority, or minorities, but they did create fear and consternation in communities where cross-burning is used as a method of intimidation and often as a precursor to violence. If posted on social media sites, the cross-burning images may have fallen foul of mainstream social media platforms, as the group has previously self-identified as supporters of National Socialism (Nazism) which is a racist and genocidal ideology. In addition, cross-burning is a racist and murderous act, particularly in the US. Yet mainstream platform policies are ambiguous in terms of how they assess actors where the group does not self-identify as white nationalist/supremacist or does not explicitly praise or support that ideology. For example, some of our members advise that a channel that promotes demographic invasion theory (used by Tarrant, Breivik, Crusius and others to justify terrorism) through continually posting third party links that falsely contextualise current events, or implicitly through memes and coded language, is quite likely to be assessed as consistent with community standards on Facebook. The platform's decision-making criteria for dangerous or hate organisations is ambiguous. Although NSN has never had a Facebook account, it has had two Twitter accounts, both of which have been removed. Instead, NSN has accounts on extremist sites.

Whilst social media is a critical tool in the early stages of the radicalisation process, its impact varies. Social media allows for rapid and widespread dissemination of information, but the terms of service and review process limits its reliability and effectiveness for extremists. Mainstream social media is a useful tool to get mainstream messages to the masses and coordinate large scale activities, like protests, but its effectiveness for extremists is limited as

it is insecure for communication and generally monitored by both social media companies and law enforcement/security services. Mainstream social media, like Facebook, Twitter, Instagram, etc. do have their use for extremists through the easy dissemination of introductory information. This information is designed to make individuals aware of ideology and group existence as well as act as a contact point. Extremists achieve this through the use of memes, short posts aimed at identifying with susceptible people. The posts are generally professionally designed and are intended to draw attention. Even when social media companies delete posts or ban accounts, posts have been copied and reposted quicker than social media companies can take them down.

Research from Macquarie and Victoria Universities also shows that content that dehumanises 'outgroups' such as Muslims, Jews and immigrants finds ways to survive on mainstream social media.[1] Some of our members provide insights, mainly through detailed observations and analysis they carry out on mainstream and extremist sites, that the more dangerous aspect of social media is the 'echo chamber' effect. Individuals who have grievances, post their grievance to social media which is then echoed back to the original poster. The feedback loop has the ability to self-radicalise an individual as they perceive the feedback as new information supporting their views, when in reality it is the same information in a different form. 'Echo chambers' are utilised by extremists to subtly recruit through this self-radicalisation process. Yet, issues with technical access to platforms such as Facebook discourage large scale analysis of these echo chambers.[2]

Within echo chambers on mainstream social media, accounts may use tactics such as disinformation to promote conspiracy theory. Techniques of 'threat construction', 'dehumanisation' and 'guilt attribution' are used to publish stories that attribute heinous crimes and offensive conduct to particular 'outgroups'. These are prominent ways in which an online community can be cultivated to see an 'outgroup' as an existential threat without explicit incitement of violence or genocide. The comment threads often include explicitly dehumanising slurs, glorification of death and genocide of particular groups, and threats of violence that are not detected by social media auto-detection frameworks, thus allowing individuals to be socialised in dangerous ways.[3] With greater freedom on low moderation sites, the discussion escalates to more explicit extremist ideology and violence.

---

[1] Department of Security Studies and Criminology. (2020, October 9). Mapping Networks and Narratives of Online Right-Wing Extremists in New South Wales (Version 1.0.1). Sydney: Macquarie University.

[2] Vermeulen, M. (2020, November 27). The keys to the kingdom. Overcoming GDPR-concerns to unlock access to platform data for independent researchers. https://doi.org/10.31219/osf.io/vnswz

[3] "In the posts, words like Jew, Jewish or Zionist appeared relatively rarely, however they were much proportionally frequent in the comments (+193%). This divergence was initially surprising given that the groups were openly anti-Semitic and that posts usually only attracted a small number of comments. There are at least two possible explanations for this. First, the group leaders were implicitly alluding to anti-Semitic tropes in some of their posts (possibly in an attempt not to breach FB community standards), which followers then made explicit. Second, group leaders simply avoided anti-Semitic messaging (possibly to avoid being shut down by FB), but the followers still brought their anti-Semitic themes into the discussion through comment.": Mario Peucker, Debra Smith and Muhammad Iqbal, 'Mapping Networks and Narratives of Far-Right Movements in Victoria' (Project Report, Institute for Sustainable Industries and Liveable Cities, Victoria University, November 2018), 9.

In recent years, social media companies have committed considerable resources to stopping social media from being used for the purposes of extremists. Despite their efforts, it has generally been too little too late. Nonetheless, extremists have seen the limitation and moved to other platforms that offer a more secure communication environment, greater anonymity, and reduction in the effectiveness of law enforcement monitoring. In other words, many right-wing extremists have moved away from mainstream social media to platforms that still allow for communication but provide better security. These new platforms include gaming platforms. Gaming platforms offer a security feature that is difficult for law enforcement to penetrate, that is, the need to play the game. For example, Minecraft is a popular game especially for younger children, but this seemingly innocuous game has a darker side, with a community of right-wing extremists utilising the platform to communicate and spread their ideology to the target audience, young children.

The dark web offers the same opportunities as secure communication technology. The ability for anonymity, secure communication, and global access. Although recent achievements by law enforcement have reduced the anonymity once offered by the dark web, it is still a viable option. One additional benefit of the dark web is the ability to access resources with relative anonymity. The availability of resources on the dark web allows for individuals to obtain resources with much less risk of detection and/or reporting than purchasing resources from mainstream sources.

Identical observations made in Australian Muslim Advocacy Network, "The Extreme Right Actors targeting the Islamic Community: Report provided to Facebook and Twitter", 24 August 2020.

See also, Emma Fell, Ellyse Anderson and Emily Hazzard, 'Like Love Project: Vilification Protections for LGBTIQ People', UQ Pro Bono Centre, November 2017, p 5: 'Hence, these comments sections tend to spiral into increasingly hateful debates. The group dynamic enhances the willingness of some people to speak up, as they are encouraged by other users' comments that align with their own attitudes.'

WAY FORWARD

Adjusting the policy frame of Online Safety

It is an uncomfortable fact that history has shown that legislative responses do not keep up with advancements in threats or technology. This needs to change; responses must keep up to date with the dynamic world of violent extremism, wherever it manifests on the violence spectrum (hate incident, hate crime or terrorist act).

Current policy definitions of 'violent extremist' or 'terrorist content' are limited by an Australian legal framework that does not recognise extremist movements, unless there is evidence of organisational structure and explicit calls to enact violence against Australians. Nor does current policy recognise extremist ideology, unless one is referring to material that urges violence or supports advocacy for terrorism. Given the points raised above, Australian policy and law must move beyond policy absolutism, and look to a range of proportionate levers to proscribe and contain (avoid the spread of) unacceptable behaviour, whilst also engendering real 'safety by design' by digital platforms. This response must be informed and directed to the full range of contributing factors to violent extremism.

For the purposes of the Online Safety Act, one option is to define violent extremism to include 'the violent denial of diversity'[4], a definition used by the Khalifer Ihler Global Institute:

> Unifying all violent extremists, regardless of their beliefs or ideological objectives is their beliefs that peaceful coexistence with someone different from them is impossible, and that violently enforcing this either through forced submission or through eradication of diversity is the solution.

This definition extends beyond terrorist violence, to capture the expression of ideology, albeit within a prism that is connected to violence. It also appropriately captures the continuum of violence that is the reality many communities experience, from online vilification and incitement, to offline hate incidents (subcrimes) and hate crimes. Not only terrorism.

Currently there appears to be a division of responsibility that places vilification and incitement to violence in the remit of the Attorney-General, via anti-discrimination frameworks, while hate crime is handled by state and territory jurisdictions, and countering violent extremism,

---

[4] 'White supremacists wish for all people in their community to be white, they thus see it as legitimate to both kill and intimidate those who are not, or those who support the idea of peaceful coexistence. Islamist extremists such as ISIS wishes for everyone to follow the same interpretation of Islam, they enforce this by killing and intimidating those they disagree with, and through threats and intimidation seek to convert others to subscribe to their interpretation. The violent components of violent extremism can broadly be categorised under the definitions of violence developed by peace-researcher Johan Galtung. He divided violence into the categories of Cultural Violence, Structural Violence and Direct violence which is a fitting framework also for understanding the violence of violent extremists': Khalifa Ihler Institute, 'Hate Map: Definitions, Scope, Terms', < https://www.khalifaihler.org/hate-map>.

The ACHN notes that the causes, origins and objectives of both those movements are more complex that this describes but acknowledges this a useful illustrative point by the Khalifa Ihler Institute about a feature that is common between the two movements.

including the early stages of extremism, resides with the Department of Home Affairs. This fragmented approach most likely inhibits cohesive policy development in the online sphere.

RECOMMENDATION 1

Expand the remit of Online Safety to include a broader concept of violent extremism that includes the violent denial of diversity as defined in this submission, to bring a cohesive policy approach to this public harm.

Civil vilification remedies through Anti-discrimination frameworks

While the AHCN recognises the important presence of civil vilification laws in a number of states and the ACT, and section 18C of the Commonwealth *Racial Discrimination Act 1975*, there are many limitations to the civil process that impede justice. These laws require individual community members to bear the personal and financial cost of seeking to have the material removed or holding an individual account, should the matter move to Tribunal or the Federal Court. The process can take months if not years. The conciliation process is impractical and unlikely to be effective in cases where respondents hold extremist beliefs. The conciliation meeting is voluntary, and most extremists or those radicalising are highly unlikely to attend a meeting. Moreover, the confidential nature of the conciliation process has long attracted criticism for privatising, and thereby burying, unacceptable behaviour that is essentially a public harm. It is designed more for complaints against work colleagues, neighbours, public figures, not people who are inciting hatred online, especially where they are espousing conspiracy theories associated with terrorists.

RECOMMENDATION 2

Recognise that existing civil vilification laws have limited effectiveness and appropriateness for online vilification, especially where an individual is espousing conspiracy theories that are associated with violent or non-violent extremist movements.

The Online Safety Act as an appropriate framework

In 2019, the e-Safety Commissioner's office undertook research in partnership with Europe and New Zealand to understand the impacts of online hate speech. From the 3700+ Australians surveyed, the overwhelming majority supported action to check the spread of online hate speech, including the introduction of legislation and getting social media companies to do more.[5]

In a submission on the second draft of the Religious Discrimination bill, more than 160 Muslim organisations argued anti-Muslim hate networks are growing online, "thanks to an environment of legal uncertainty". Referring to the March 15 2019 shooting that killed 51 people, the Muslim community said: "The atrocity of the Christchurch terror attack continues to reverberate in the Australian Muslim community," and "Australian Muslims have never felt this unsafe."[6]

Prominent Indigenous identities contributed to an open letter to Attorney-General Christian Porter demanding stronger penalties for perpetrators of online racism. The letter requests "urgent, further action" to make online racists legally accountable, as well as permanent bans for individuals who post racist material on social media and a national action plan to respond to racist comments online.[7]

Over many years, Australia's Jewish Community have also been calling for more effective vilification and incitement laws.[8]

A report considering the effectiveness of Queensland's vilification laws to target anti-LGBTIQ hatred, has pointed to Facebook as providing a 'disinhibiting environment' for people to

---

[5] Australian Government, e-Safety Commissioner, *Online Hate Speech: Findings from Australia, New Zealand and Europe* (January 2020), < https://www.esafety.gov.au/about-us/research/online-hate-speech >.

[6] Judith Ireland, ''Never felt this unsafe': Muslim community pleads for more protection in religious discrimination bill', Sydney Morning Herald, 8 March 2020.
See also, Zena Chemas, 'Victims of racist attacks fear reoccurrence as Australia launches inquiry into right-wing extremism', ABC News, 19 December 2020, < https://www.abc.net.au/news/2020-12-19/muslim-women-speak-out-on-trauma-amid-right-wing-inquiry/12961164 >.
Australian Muslim Advocacy Network, Submission to the UN Special Rapporteur's Report on Anti-Muslim Hatred and Discrimination < http://www.aman.net.au/wp-content/uploads/2020/12/UN-Special-Rapporteur-Submission- -Aust-Muslim-Advocacy-Network.pdf>

[7] Mikele Syron, 'Open letter demands 'urgent, further action' to stop online racial attacks', NITV News, 9 September 2020. To see the Open Letter spearheaded by Shelley Ware and Professor Kate Seear, go to https://www.outersanctum.com.au/updates/2020/9/1/open-letter-to-the-commonwealth-attorney-general.

[8] See recent example, Linda Mottram, "As the Holocaust is remembered, a call for Australia to ensure effective laws against vilification and incitement", ABC Radio PM interview featuring Julie Nathan, Research Director of the Executive Council of Australian Jewry, 27 January 2021.
https://www.abc.net.au/radio/programs/pm/warning-anti-semitism-continues-to-increase-in-australia/13096500

display their most hateful sides. The report found the effectiveness of safeguards proposed by Facebook to be 'unrealistic', 'largely inadequate' and unsafe to users[9].

The proposed Online Safety Bill 2020 drives 'safety by design' for digital platforms by:

(1) establishing an accountability framework through data, reporting and priorities that need to be reported on (through the "Basic Online Safety Expectations");

(2) promoting best practice in regard to material that is in the Online Content Scheme (through "Industry Standards" and "Industry Codes"); and

(3) providing accountability through complaint processes and consequences, including civil penalty provisions (there are 5 different schemes).

Therefore, this Bill aims to provide a feedback loop for continual improvement at a design level for digital platforms; along with some redress for victims through civil penalty provisions and other remedies.

RECOMMENDATION 3

Recognise that the Online Safety Act provides an optimum and comprehensive framework to push back against online echo chambers that socialise individuals towards accepting the violent denial of diversity, as well as individuals who serially post vilifying or violence-inciting material.

[9] See Fell et al, above n 3, 7: 'These mechanisms are largely ineffective at preventing, or remedying, unlawful vilification…unfriending or blocking the individual responsible for the offending content fails to understand the public nature of the platform, as 'friendship status' is not required to view another person's comments. In addition, the suggestion that the victim unfollow or block the public page has the consequence of inappropriately punishing the victim for the perpetrator's offensive conduct. Lastly, despite its suggestion that victims report abusive content, it has been reported that Facebook is failing to remove vilification on the grounds of LGBTIQ status.'

Mitigating incitement to violence, glorification of genocide and incitement to commit genocide

The proposed Online Safety Bill proscribes material that *promotes* abhorrent violent conduct, material that *incites* abhorrent violent conduct, material that *instructs* in abhorrent violent conduct or abhorrent violent material. Packaged together within the 'Abhorrent Violent Material Scheme', it is intended to be invoked when there is a crisis event. It defines abhorrent violent conduct in accordance with the Commonwealth Criminal Code:

Criminal Code 474.32 Abhorrent violent conduct

(1) For the purposes of this Subdivision, a person engages in *abhorrent violent conduct* if the person:
(a) engages in a terrorist act; or
(b) murders another person; or
(c) attempts to murder another person; or
(d) tortures another person; or
(e) rapes another person; or
(f) kidnaps another person.

This is enacted through an Abhorrent Violent Material (AVM) scheme which includes a process for complaints and civil redress, along with the options for take-down requests, link deletion requests and blocking requests. The Basic Online Safety Expectations also require platforms to minimise this content, and provide periodic reporting.

Therefore, inciting or promoting terror, murder, torture, rape, kidnapping against another person is a harm that can trigger civil penalties via the Online Safety Bill, where in connection to a crisis event. However, if a person urges violence or harm generally against a person or classes of persons, it is unlikely to meet this threshold. Currently, the scheme is not intended as a general incitement to violence clause. Inciting or promoting genocide is also not listed as Abhorrent Violent Conduct.

There is potentially limited scope under the Online Content Scheme to regulate incitement to violence, as it gives the e-Safety Commissioner scope to take down material that appears to be RC Classified. The description of content that is Refused Classification (RC) under the National Classification Code includes materials that "depict, express or otherwise deal with matters of … crime, cruelty, violence or revolting or abhorrent phenomena in such a way that they offend against the standards of morality, decency and propriety generally accepted by reasonable adults to the extent that they should not be classified". Whether this is intended to include incitement to violence or incitement to genocide is questionable. In any event, explicit inclusion of these concepts is preferable. Otherwise, the Australian community may not know there is scope to take immediate action, and the data secured on incitement to violence is likely to be ad hoc.

Some might argue that incitement to violence is best managed by the criminal system, but this does not provide for the fast removal of content, or penalties to platforms – and criminal incitement to violence laws have been very poorly enforced in Australia. The only federal

incitement to violence law has never been used to successfully convict, with only one commenced prosecution. [10]

Some of our members report that examples of online incitement to violence identified by communities have not been pursued by police. This may be due to the scale of online incitement, but police prosecutors or DPP/CDPP tend to look for evidence of more serious conduct in tandem with incitement – such as providing instructions to carry out a violent attack, or additionally, planning an attack. This is not something our members necessarily agree with and we will be making recommendations in other forums about the way criminal laws are implemented. However, given the platform-facing framework of the OSA, it is critical that civil mechanisms also be animated.

RECOMMENDATION 4

Recognise in relation to online incitement to violence and murder, including genocide, and glorification of violence, murder and genocide, that civil mechanisms integrated within a 'Safety by Design' policy framework are necessary to complement the criminal system and should be defined in a common-sense manner. This should include:

- **fantasising that an identified group was dead, would die, or be left for dead;**
- **suggesting that an identified group should be 'dealt with' in the way that this identified group, or another identified group, was 'dealt with' in a former genocide;**
- **celebrating or glorifying genocide or an atrocity where members of an identified group have died;**
- **wishing that more members of an identified group died in a genocide or atrocity;**
- **suggesting, implicitly or explicitly, that further violent attacks should take place against an identified group;**
- **suggesting that violence carried out by an individual be returned to all members of that individual's identified group or that all members of that group should have to pay the price; or**
- **calling for violence or a form of violence against an identified group or members thereof.**

'Identified group' refers to a group identified, either explicitly or clearly implied, on the basis of race, religion, gender, gender identity, sexuality or disability status.

[10] Disclosed via the Attorney-General's response to a Question on Notice from Senator Kristina Keneally in 2020.

The goal is to push back against online echo chambers that socialise individuals towards accepting the violent denial of diversity.

RECOMMENDATION 5

Include definitions for vilification, incitement to violence, glorification or incitement to genocide within the Online Safety Act, consistent with existing Australian law, so they may become:

(1) **Matters of complaint to the e-Safety Commissioner, resulting in the possibility of more effective, quick take-downs of material**

(2) **Priorities within the Basic Online Safety Expectations, prompting greater accountability and the creation of longer term data through periodic reporting by platforms**

(3) **The subject of an Industry Code or Standards, if included within the Online Content Scheme.**

Strategic action against vilification

At an operational level the e-Safety Commissioner could prioritise accounts or administrators who serially produce or publish such material, rather than individual posts or comments made on comment threads. Strategically, one of the biggest challenges is lifting platform performance when it comes to enforcing Australian standards on vilification and incitement to violence. With a very heavy reliance on auto-detection frameworks, the ability of mainstream platforms to escalate, competently and consistently assess online accounts or administrators who serially post vilification or incitement to violence within an echo chamber, is weak. It needs improving through Industry standards for how to competently and consistently assess actors, using universal markers that are not ideology-dependent or constrained by government's terrorist proscription lists. Standards have the ability to articulate contextual factors that ought to be considered in assessing a particular actor, which could also be used by the office of the e-Safety Commissioner. Platform performance can also be improved through civil penalties for non-compliance, which will influence their resource prioritisation. The degree to which auto detection frameworks can competently assess an actor's behaviour *in context* is still very problematic, and human expertise needs to be better resourced by platforms. We also need to better understand the scale of the problem through data.

RECOMMENDATION 6

Expand the remit of the Online Content Scheme so that the e-Safety Commissioner can establish Industry Standards in relation to assessing online actors that serially publish vilifying or violence-inciting material to 'echo chamber' and other audiences.

The role of vilification in providing a license to commit violence

A usefully distinct concept within vilification is the subset harm of dehumanisation. It is raised here as Australian researchers[11] have begun to point to dehumanisation as a central tenet of violent extremism. While ideology may mix with other personal factors to trigger violence, one characteristic of ideologies across the spectrum is dehumanisation of the 'other' or designated 'outgroup'. Vilification more broadly can embolden individuals to consider carrying out hate incidents, hate crime, and lower their resilience to violent extremist materials. Dehumanising the target group helps perpetrators to overcome any moral or other objection to committing murder or other serious harm, by seeing the victim group as a pest that needs to be exterminated, a threat to the existence or health of the dominant group.

Dehumanisation is also an established legal concept within international law.[12] Below are some recent and historical examples that briefly touch on the connection between dehumanisation and genocide:

Armenian case Historically, Armenians were considered a fifth column, or traitors, and regularly dehumanised as germs, bacteria, microbes that needed to be eradicated. [13] One academic, Holslag, notes the "symbolism at work in the selling of female Armenians: the symbolism of the dehumanization and commercialization of the victimized group… By commercializing victims and using them literally as a product, perpetrators transformed victims from humans into non-humans." Even today, Armenians who survived the genocide are pejoratively called "remnants of the sword".
Nazi era There is prolific evidence on how Jews were dehumanised during and in the lead up to the Holocaust, both with words like vermin, parasites, and with related images. Similarly, other groups the Nazis targeted such as people with disabilities, and Roma/Sinti, were also described in derogatory ways (eg, unworthy of life; criminals etc).[14]

[11] Above n 1.

[12] See discussion in H J Van Der Merwe, 'The Prosecution of Incitement to Genocide in South Africa' (2013) 16(5) Potchefstroom Electronic Law Journal 327.

[13] See for example, the notes of a practising doctor in: Robert Kaplan, 'Monstrous Complicity: Doctors and the Armenian Genocide', *ABC Religion and Ethics*, 21 April 2015.

[14] Johannes Steizinger (2018) The Significance of Dehumanization: Nazi Ideology and Its Psychological Consequences, Politics, Religion & Ideology, 19:2, 139-157; Bain, P.G., Vaes, J., & Leyens, J.P. (Eds.). (2013). Humanness and Dehumanization (1st ed.). Psychology Press.

Cambodia

During the Cambodian genocide, enemies were described as 'worms' (*dangkow*) who 'gnawed the bowels from within' (*see roong ptai knong*). They represented 'no loss' (*meun khaat*) when 'weeded out,' (*daak jenh*) whilst victims of enforced migration were 'parasites' (*bunhyaou k'aek*) who 'brought nothing but bladders full of urine' (*yoak avey moak graowee bpee bpoah deuk*). The sick were 'victims of their own imagination,' (*chue sotd aarumn*) unlike the party who remained 'strong' (*kleyang*) and 'healthy' (*dungkoh*).[15]

Rwanda

Dehumanisation was a key strategy in persecuting Tutsis and in inciting violence by Hutus. There is a lot of information on this, especially because it was broadcast over the radio.[16]

Bosnians

Refic Hodzic writes: 'Dehumanising Bosnian Muslims to filth, spawn of Ottoman occupiers, traitors who sold the ancestral faith, vermin that needs to be annihilated and removed is Karadžić's most devastating legacy. The degree of stripping Muslims of any human value is best illustrated in the statement of Biljana Plavsić, Karadžić's deputy and right hand, referring to Bosnian Muslims:

"It was [Serb] genetically deformed material that embraced Islam."
Such characterizations ultimately enabled unimaginable cruelty to be unleashed by neighbours on neighbours.

This is what normalized the killing of 102 children in my hometown, for example. I know who these child-killers are, I see some of them on the streets of Prijedor. It is utterly wrong to think of them as deranged psychopaths or crazies who were somehow driven insane by hatred, they were ordinary people, yesterday's best friends, workmates, childhood buddies of their victims (Slavenka Drakulic captured this dimension well in "They would not hurt a fly").

The dehumanisation of Muslims implemented in a concerted effort by the Serb elite – political, religious, intellectual – and disseminated by the media conditioned the people who were to do the killing to see it as a necessary, dirty work that simply must be done if Serbs are to enjoy freedom free of "invaders".[17]

[15] Fionn Travers-Smith, 'How the Khmer Rouge Dehumanised their "enemies"', New Mandela, 11 June 2012 < https://www.newmandala.org/how-the-khmer-rouge-dehumanised-their-enemies/>

[16] See Kennedy Ndahiro, In Rwanda, We Know All About Dehumanizing Language
Years of cultivated hatred led to death on a horrifying scale, *The Atlantic*, 31 April 2019.
https://www.theatlantic.com/ideas/archive/2019/04/rwanda-shows-how-hateful-speech-leads-violence/587041/

[17] Refic Hodzic, 'Dehumanisation of Muslims made Karadzic an icon of far-right extremism', JusticeHub, 22 March 2019 < https://justicehub.org/article/dehumanisation-muslims-made-karadzic-icon-far-right-extremism/>.

Yazidi examples[18] (often overlapped with misogyny)
ISIS is a terrorist organisation that propagates an ideology grounded in a perverted interpretation of Islam that has been condemned by Muslims worldwide.

Yazidis were considered by ISIS as 'mushrik' or idolaters. They were denigrated as devil-worshippers and polygamists (they are actually monotheists).

ISIS also saw Yazidi women as vessels to impregnate and make Muslim babies. Many Yazidi women and girls were stripped naked, numbered and offered for sale online or in marketplaces. There was a system for inventorying them, referring to them as sabaya (slave) followed often by a number. Numbering people is a form of removing their identity and dehumanising them (as also seen during the Holocaust).

According to first hand reports, one ISIS captor sat on a pregnant woman's stomach trying to kill her unborn baby (she was pregnant when she was kidnapped) saying "this baby should die because it is an infidel; I can make a Muslim baby."

Other words that reduced Yadizi people to things were commonly used, such as 'share' "You can sell your slave, or give her as a gift… You can do whatever you want with your share."

The label of infidel and 'kuffar' is linked to violence during the genocide of Yazidis, eg "He told me that according to Islam, he is allowed to rape an unbeliever. He said that by raping me, he is drawing closer to God."

One supporter wrote on social media: "Yes they are idolaters, so it's normal that they are slaves, in Mosul they are closed in a room and cry, and one of them committed suicide LOL [laugh out loud]." [19]

Rohingyas in Myanmar
The report of the Fact Finding Mission on Myanmar, sets out in extensive detail its findings on the extreme violence perpetrated against the Rohingya in Rakhine State since 25 August 2017, in what the Tatmadaw referred to as 'clearance operations'.

It documents in unsparing detail how the Tatmadaw took the lead in killing thousands of Rohingya civilians, as well as forced disappearances, mass gang rape and the burning of hundreds of villages.

[18] Marczak N. (2018) A Century Apart: The Genocidal Enslavement of Armenian and Yazidi Women. In: Connellan M., Fröhlich C. (eds) A Gendered Lens for Genocide Prevention. Rethinking Political Violence. Palgrave Macmillan, London. https://doi.org/10.1057/978-1-137-60117-9_7

[19] This is also seen during the Holocaust, where human bodies were referred to as "stück" (piece) – many other examples of this in the Nazi case.

The report also explains how hate speech and false narratives were spread on Facebook and other means, including by military and government officials. It portrayed the Rohingya as an invasion, and as savage criminals that threatened the existence of other races. This shows how dehumanisation is also achieved cumulatively through discourse, especially disinformation. There were also examples of more explicit dehumanisation: "watering poisonous plants."[20]

Many of the derogatory terms were not picked up by moderators on Facebook.

Authorities refuse to acknowledge their identity as Rohingya, rather insisting they are Bengali, and relatedly, that they have illegally migrated to Myanmar. Denial of identity can be linked to dehumanisation, though this seems more a way to construct them as Other and as a community that doesn't belong.[21]

Uighur women have been called 'baby making machines' by the Chinese Government, however this prompted action by Twitter.[22]

Dehumanisation is increasingly recognised as a technique used in propaganda of extremist movements and terrorism. For example: in the manosphere/INCEL movement[23]; in relation to the 'counter jihad' movement[24] and Tarrant's manifesto;[25] in relation to far right movements more generally;[26] in Islamist movements such as ISIL; and often in ultra-

[20] United Nations Human Rights Council, The report of the Fact Finding Mission on Myanmar, 18 September 2018, p 166. <https://www.ohchr.org/EN/HRBodies/HRC/Pages/NewsDetail.aspx?NewsID=23575&LangID=E>

[21] That is referred to in passing here: https://www.abc.net.au/religion/the-rohingya-genocide-does-not-end-at-myanmars-borders/10095384; https://thediplomat.com/2019/12/myanmars-rohingya-vs-bengali-hate-speech-debate/

[22] "Chinese Embassy Twitter account locked for "dehumanization"', https://apnews.com/article/race-and-ethnicity-china-c3175de6dfd85ad019f05b59dc26d22f.

[23] Sian Tomkinson, Katie Attwell, Tauel Harper, ''Incel' violence is a form of extremism. It's time we treated it as a security threat,' *The Conversation,* 27 May 2020 < https://theconversation.com/incel-violence-is-a-form-of-extremism-its-time-we-treated-it-as-a-security-threat-138536 >.

[24] Rita Jabri-Markwell, "The online dehumanisation of Muslims made the Christchurch massacre possible," ABC, Aug. 31, 2020.

[25]'Tarrant indicated that, when trying to remove a nest of snakes, the young ones had to be eradicated. Regrettably, children were among those whom he allegedly shot and killed': see Peter Lentini, 'The Australian Far-Right: An International Comparison of Fringe and Conventional Politics' in Mario Peucker and Debra Smith (eds), *The Far-Right in Contemporary Australia* (Singapore, 2019) 19, 43. An analysis of the link between Tarrant's terror attack and anti-Muslim dehumanisation in the Bosnian genocide, see above n 6.

[26] See Simon Copland, 'How do you prevent extremism?', BBC In Depth Psychology, 2 May 2019; Wahlström M, Törnberg A, Ekbrand H. Dynamics of violent and dehumanizing rhetoric in far-right social media. *New Media & Society*. August 2020.

nationalist movements. Although not intentionally evaluated, dehumanisation is also evident within Australian research on the way that the far-right milieu characterises Muslims, Jews, Indigenous Australians, people of African and Asian backgrounds, LGBTIQ people and women.[27] This illustrates why addressing incitement to violence alone would not be enough to mitigate against online echo chambers that socialise individuals towards the violent extremism.

A significant problem now is the way that disinformation and dehumanisation are coming together. Over time, many stories may be published about a victim group, in order to attribute collective guilt to their identity for heinous crimes and savagery (subhumanity), or to present them as a homogenous hostile mass that does not independently think or feel (ie mechanically inhuman), or as demonic form that is orchestrating the world's harms or downfall of society (ie supernaturally inhuman). However, this conduct does not trigger the dehumanisation policies of mainstream social media. Their policies are more oriented towards explicitly dehumanising language, something that can be regulated via automated systems. Contemplating the more cumulative effects of disinformation and dehumanisation via discourse requires consideration of context, a task better suited to human expertise, which platforms resist due to resourcing.

RECOMMENDATION 7

Consider dehumanisation and disinformation (via false narratives) as usefully distinct concepts within the broader category of vilification, when developing policy and Industry Standards to assess extremist behaviour and echo chambers online.

[27] Peucker, Smith and Iqbal, above n 3.

Profiting from vilification and incitement to violence online

Websites from actors within extremist movements are often hosted overseas. For example, neo-Nazi websites by Australians are generally hosted in the US. Therefore, it is extremely difficult to get them taken down.

Additionally, mainstream retailing websites have been caught out selling hate products, but it is only with considerable media attention and public pressure that action is taken. That action is not consistent across the board as there are no clear legal obligations or responsibilities. For example, in 2019, an online retailer named Redbubble was selling Auschwitz themed clothing and tote bags, that featured images of Auschwitz.

The Institute of Strategic Dialogue also recently published a ground-breaking report on the use of mainstream fundraising tools to raise funds for hate-based movements in the US. It found anti-LBGTIQ and anti-Muslim hate-based organisations enjoy almost unfettered access to mainstream online fundraising tools in the United States with a less substantial, but still concerning amount of white supremacist organisations still being able to fundraise.[28]

RECOMMENDATION 8

Host providers, online retailers and fundraising platforms must safeguard against the risk of actors being able to financially profit through the vilification of protected groups or the promotion of dehumanising or violent ideology. The e-Safety Commissioner needs to be able to request an Industry Code or establish an Industry Standard to articulate these specific obligations.

[28] Institute for Strategic Dialogue and Global Disinformation Index (2020) *Bankrolling Bigotry: An overview of the online funding strategies of American hate groups.*

The liability of third parties who fail to moderate online

A 2019 NSW Court of Appeal decision had wide-ranging ramifications for media companies in Australia, which are now held responsible for defamatory content posted by users on their Facebook pages.[29] Multiple communities represented within the AHCN advise that certain organisations, agencies and especially news services, do not take their responsibilities for moderating content seriously, allowing vilifying, violence-inciting and even genocide-glorifying language to proliferate in comment threads on material they post. In regard to news services, sensationalist headlines, with articles even at times hidden behind paywalls, are particularly problematic. Articles that touch on sensitive issues also require particular attention by moderators. This Act should consider ways to incentivise those third-party organisations to share a serious sense of responsibility with platforms to moderate content.

RECOMMENDATION 9

Within the Online Safety Act, clarify the liability of third parties, especially news services, to moderate abusive, vilifying or inciteful content in the comment threads of articles they post.

[29] Brett Walker, 'Voller defamation case highlights law's struggle to keep pace in digital age, says ANU Law expert', ANU College of Law, 11 July 2019 < https://law.anu.edu.au/news-and-events/news/voller-defamation-case-highlights-law's-struggle-keep-pace-digital-age-says-anu >.

Competent decision-making

Further consultation with targeted communities to understand specific harmful, online practices and speech with regard to their context is needed. For example, in terms of online safety for LGBTIQ+ people, one of the most dangerous and harmful practices is "outing" someone. This is not always detected as bullying, abuse, or even 'hate speech', but can result in violence, intrafamilial hate crime, and in some extreme circumstances, homicide (see for example the recent murders in Russia stemming from a campaign of outing LGBTIQ+ people). Similarly, for disabled people, speech often deployed in argument – such as stupid, idiot, moron – is considered ableist hate speech by some disabled people. Additionally, in terms of online safety, for some disabled people (with cognitive impairments), the main issue is the manipulation of their vulnerability – often from "mates" (see the term "mate crime"). This also relates to the experiences of older people, for whom elder abuse may present as hate crime. Certain groups are harmfully characterised via conspiracy theory and disinformation, which also requires awareness from decision-makers who are interpreting the Act. Without further engagement with other targeted groups, these unique and poorly understood forms of online abuse and harassment may be missed.

RECOMMENDATION 10

Consult further with targeted communities to understand specific harmful, online practices and speech with regard to their context, with a view to creating decision-making guidelines for those administering the Act.

CONCLUSION

The proposed Act deals with a scenario where a child or adult is personally and viciously abused – but remains silent on actors who serially post vilifying material, or the echo chambers that socialise individuals towards the violent denial of diversity.

This Act has potential because it carries a range of levers and sets expectations for both online providers and the online community of Australian users. It is also about gathering data to hold platforms to account. Our concern is that its attention to online hatred is inexplicably limited, at a time when platforms need to be encouraged in every way possible to dedicate more resources to properly assessing groups, channels, pages and accounts that serially vilify, incite violence or glorify genocide. The elements of the Act that use accountability, data reporting and standards to push constant improvement by platforms must be explicitly configured to enable quality policy development in this area.

In regard to the complaint mechanisms, the imperatives of ensuring administrative officers are only making decisions with regard to more serious harms can be managed through focusing on particular strategic focal points or 'vectors', like those users who serially post vilifying or violence-inciting material, or those echo chambers which socialise individuals towards the violent denial of diversity.

The Australian Government is urged to hear the call from Australian people and strengthen this Act to address the full continuum of violence endangering the safety of Australian users.

We appreciate your consideration of these concerns and recommendations.