ALICE LIDDELL

REMOVING OBJECTIONABLE CONTENT FROM ONLINE PLATFORMS IN NEW ZEALAND: CAN TAKE-DOWN NOTICES RISE UP TO THE CHALLENGE?

LLM RESEARCH PAPER

LAWS 582: LEGAL WRITING

FACULTY OF LAW

VICTORIA UNIVERSITY OF WELLINGTON
TE HERENGA WAKA

2021

Contents

Abstract

The Films, Videos and Publications Classification Act 1993 is likely to be amended in 2021 with the aim to prevent and reduce the harm caused by objectionable publications. This paper considers take-down notices alongside theories and international examples of regulating content on social media. Governments and tech companies cooperate to combat extremist content online, particularly terrorism and child exploitation. Governments previously had limited ability to enforce domestic laws against global social media platforms when the limits of what can be achieved with cooperation are reached. Take-down notices, and removal of 'safe harbour' for content hosts are considered in terms of enforceability and effectiveness.

Word length

The text of this paper (excluding table of contents, abstract, footnotes and bibliography) comprises approximately 11,998 words.

Subjects and Topics
objectionable publications-social media-online regulation-take down powers

I Summary

Governments are now attempting to curtail the power of social media platforms through passing laws placing responsibility on platforms to prevent, minimise and remove illegal content. This paper is focussed on the potential for New Zealand to make and enforce laws to require the removal of objectionable content hosted on social media platform operators whose operations are largely based overseas. New Zealand is making changes to the Films, Videos and Publications Classifications Act 1993 (FVPC Act) to fill in some gaps in the legislation identified after the terrorist attacks in Christchurch on 15 March 2019. The upcoming changes will introduce a power to issue take-down notices for objectionable content published online. Take-down powers are necessary but will not drastically alter the online environment because platforms are currently voluntarily removing objectionable content when asked to by authorities. New Zealand is so far taking a cautious approach by not placing a duty of care on platforms to take ownership and minimise harmful content. There is a large volume of commentary on the risks of regulating online platforms but very little evidence of what works. By not taking an immediate reactive approach, New Zealand has the opportunity to wait for evidence to emerge before considering whether stronger laws are appropriate.

Some of the theories of how to regulate multinational social media platforms are explored. New Zealand's impending changes to the FVPC Act are discussed and compared with two overseas examples, Germany and Australia. Since 15 March 2019, New Zealand has asserted itself in a position of leadership with the development of the Christchurch Call. There is momentum for asserting power over social media platforms but taking control requires a bold position and the ability, resources and will to take enforcement action. However, New Zealand is taking a cautious intervention approach so far compared with other countries which have passed laws relating to harmful online content.

Social media platforms are no longer seen as a post box or piece of infrastructure. They profit from content traffic and have huge power in controlling what content users are directed to. With this great power, some responsibility is required. While it is entirely possible to make laws targeted at global social media platform operators, it is another thing to enforce those laws and hold platforms to account for hosting illegal content. While there are challenges in enforcement and holding large internet companies to account, even small countries like New Zealand can take control and set expectations.

II Background

A Christchurch terror attacks and social media

On 15 March 2019 in Christchurch, New Zealand, a terrorist attack was carried out at two mosques, killing 51 people and injuring a further 50. The terrorist livestreamed the attack on Facebook using a head-mounted GoPro camera. The 17-minute livestream was viewed 4,000 times before being reported and removed 29 minutes after it started and 12 minutes after it had finished.[1] In the first 24 hours, Facebook removed 1.5 million videos of the attack with 1.2 million of those blocked at the upload stage. Facebook's algorithms did not initially detect the objectionable content in the livestream, so the video was able to be disseminated globally and published on other platforms such as 8chan and Reddit.[2] The video of the Christchurch terrorist attacks was not detected as there was no other videos like it and it resembled a video game.[3] The terrorist also used the large platforms Facebook and Twitter as beacons for publishing links to his manifesto which he published before the attacks on smaller file-sharing platforms.[4]

While the use of online platforms to disseminate objectionable content is not a new problem,[5] the 15 March 2019 attacks represented a turning point because the use of social media was an integral part of the attacks. The attacks were "designed to go viral[6]". New Zealand Prime Minister Jacinda Ardern stated on 19 March 2019 that while hate and division has long existed, the tools of distribution are new, and something must be done to hold platforms to account.

> We cannot simply sit back and accept that these platforms just exist and that what is said on them is not the responsibility of the place where they are published. They are the publisher. Not just the postman. There cannot be a case of all profit no responsibility.[7]

---

[1] Facebook *Update on New Zealand* (18 March 2019) <about.fb.com>.

[2] Peter A. Thompson "Beware of Geeks Bearing Gifts: Assessing the Regulatory Response to the Christchurch Call" (2019) The Political Economy of Communication 7(1), 83–104 at 83.

[3] Claire Mason and Katherine Errington *Anti-social media: Reducing the spread of harmful content on social media networks* (The Helen Clark Foundation, 2019) at 8.

[4] Tech against terrorism Analysis: New Zealand attack and the terrorist use of the internet (26, March 2019) <www.techagainstterrorism.org>.

[5] Twitter for example was used by ISIS to promote terrorist activity at a peak around 2015 before Twitter invested in anti-terrorism capabilities. Danny Yadron "Twitter deletes 125,000 Isis accounts and expands anti-terror teams" (The Guardian, 5 February 2016) <www.theguardian.com>.

[6] Christchurch Call <www.christchurchcall.com>.

[7] Rt Hon Jacinda Ardern "PM House Statement on Christchurch mosques terror attack" (19 March 2019) <www.beehive.govt.nz>.

The events of 15 March 2019 broadened the focus of what is considered to be extremist content online to also look at insider threats and far-right extremism. Commentary before 2019 was based on foreign terrorist content being disseminated on social media platforms, particularly the prolific use of social media by Islamic State supporters. It is acknowledged that the 15 March 2019 attacks are not the only example of a terrorist using social media, however it represented a catalyst for change in New Zealand and launched international conversations about the responsibility of platforms for content.

New Zealand authorities currently rely on cooperation and voluntary removal of content when objectionable publications are detected on internet platforms. The taking down of the video and manifesto by platforms was entirely voluntary. The New Zealand government had no power to compel any platform to remove it. The people who possessed or distributed the objectionable publications committed an offence,[8] but there was no offence that applied to platforms hosting the content. Platforms were compliant with government requests and there is no evidence to suggest that take-down powers would have hastened the removal process, however risks have been highlighted and there is no legal mechanism for removal if cooperation breaks down.

The events and aftermath of 15 March 2019 demonstrated that taking down objectionable content is not a simple task. Computer files have their own "fingerprint" or "hash" which can be used to remove or block other uploads of the same file however this can change if the file is resized, edited or mixed with other content. Once the fingerprint changes, a new version of a known objectionable file will not be automatically detected.[9] The footage of the 15 March 2019 attacks has emerged online in various forms.[10] Every time a new publication appears, the authorities rely on user reporting and voluntary removal by the platforms. Large platforms such as Facebook, Microsoft, Twitter and YouTube have agreements in place to share fingerprinting data of known terrorist content, however this is on a voluntary basis.[11] Within a few days of the attacks, Facebook had shared the fingerprints of more than 800 versions of the attack video with the Global Internet Forum

[8] In *R v Arps* [2019] NZDC 11547 the defendant was sentenced to 21 months' imprisonment on two charges under the Films, Videos and Publications Classification Act 1993 for distributing the video of the terrorist attack.

[9] Mason and Errington, above n 3, at 8.

[10] A recent example was an animated GIF containing screenshots of the attacks reported on Twitter. Radio New Zealand "Authorities move to take down 'atrocious' mosque attack material" (25 June 2021) <www.rnz.co.nz>.

[11] Michelle Roter "With Great Power Comes Great Responsibility: Imposing a Duty to Take down Terrorist Incitement on Social Media" (2017) 45 Hofstra Law Review 1379 at 1400.

to Counter Terrorism.[12] The rapid pace of spread across multiple platforms and the recreating of different versions demonstrates the challenging environment for enforcement and compliance. Take-notices are a small tool for a challenging environment.

B The power of social media platforms

1 A powerful tool

Social media is a powerful global tool for connecting people, spreading information and giving communities a voice. Social media platforms are multi-dimensional and multi-use because they typically offer the ability for users to view content, post and share content, comment on others' content and engage in private messaging. Platforms can be described as internet intermediaries[13] or gatekeepers that people use to interact online.

Platforms have changed the way industries operate such as advertising and print media. Social media companies use the data of users to sell targeted advertising, this advertising has moved beyond consumer products to political messaging and targeted news media. Platforms have the power to change the everyday lives of users by influencing values and beliefs.[14] The revelations about Cambridge Analytica harvesting Facebook user data in 2015-2016 was an example of the power of social media platforms to influence politics. Demographic technology designed for targeted consumer advertising was used to identify and exploit political persuasions.[15] It is difficult to fully understand, monitor and measure this power when platforms are continuously evolving in their technology and market position.[16]

2 A few platforms dominate the market

The information that that public receives is filtered through a small number of platforms operated by a smaller number of companies.[17] This makes regulation seem like a daunting

---

[12] Facebook, above n 1.

[13] Alex Rochefort "Regulating Social Media Platforms: A Comparative Policy Analysis" Communication Law and Policy, (2020) 25:2, 225-260 at 228.

[14] The Workshop *Digital Threats to Democracy* (2019), at 38

[15] Curtis R. Barnes, Tom Barraclough "Digitised lies: New Zealand and the globalised disinformation threat" in Andrew Chen (ed.) *Shouting Zeros and Ones: Digital Technology, Ethics and Policy in New Zealand* (2020, Bridget Williams Books, online ed.) at [20].

[16] Orla Lynskey *Regulating 'Platform Power'* (LSE Law, Society and Economy Working Papers 1/2017) at 7.

[17] Facebook, Facebook Messenger, Instagram and WhatsApp are all owned by Facebook. Facebook bought Instagram in 2012 and WhatsApp in 2014.

task. Platform monopolies are a core problem identified by New Zealand research organisation the Workshop. When there are only a few companies controlling the majority of communications and content distribution, the global market power is huge.[18] The concern about platform monopolies arises from the model of companies striving to be the one dominant platform in their field whether that be as a search engine, an online retailer, a video platform, content streaming service or social network. When New Zealand's Privacy Commissioner John Edwards delivered a speech to the International Association of Privacy Professionals, he talked about the asymmetries in the information received by users and the problem of digital monopolies. He phrases the problem as "they are one and we are many" when there are singular platforms providing a global service for diverse states and populations.[19] Ultimately Edwards believes New Zealand can push back and is not impotent against large platform operators despite representing a small number of consumers.[20] It is aspirational that a small country could effectively curb the power of large platform. The market dominance is unlikely to change so governments need to find a solution that platform operators will reasonably comply with.

3 Platforms are gatekeepers

For the purpose of regulation, one way to define captured platforms is by their gatekeeping characteristics. Using online platforms including social media and video sharing platforms is not essential however they provide convenience and access to information and audiences. Online gatekeepers who enable speech by providing a platform and access to an audience inevitably will facilitate the spread of illegal content so they should have a moral and social responsibility to prevent, minimise and remove it.[21] Social media platforms have terms and conditions or community standards which users agree to in return for the right to participate on the platform. If people interact and do business through a few internet companies in monopoly positions and they become banned or the online environment becomes unbearable due to harmful content, then their access to services is significantly reduced.

Sam Shead "Facebook owns the four most downloaded apps of the decade" (BBC News, 18 December 2019) <www.bbc.com>.

[18] The Workshop, above n 14, at 14.

[19] John Edwards "Dwarfed by the digital giants, here's how we can make our voice heard" (The Spinoff, 31 October 2019) <www.thespinoff.co.nz> (A transcript of a speech entitled "Addressing the Power Asymmetry of the Big Technology Companies", delivered at the International Association of Privacy Professionals conference in Sydney).

[20] Above.

[21] Raphael Cohen-Almagor "The Role of Internet Intermediaries in Tackling Terrorism Online" 86:2 Fordham Law Review 425-454 at 426.

Every-Palmer uses the concept of "digital gatekeepers[22]" which he believes should be regulated as "public utility services[23]". Lynskey expands on this idea preferring the concept of "digital gatekeepers" when discussing how to regulate platform power.[24] Lynskey has looked at various definitions for example requiring interaction or multi-sidedness where users can do more than just view content and has found definitions to be overly simplistic and not useful.[25] Lynskey believes a regulatory focus on problematic gatekeeper conduct rather than particular technology lessens the risk of targeting the wrong thing or becoming irrelevant due to technological change.[26] It is better to focus on the concerning practices that are in need of regulation than the platforms themselves as platforms are difficult to define and often changing.

4 Platforms are more than infrastructure

Large social media platforms have developed under a lack of regulation and lack of accountability for the consequences of rapid technological development. US based platforms developed with immunity from liability as carriers under section 230 of the Communications Decency Act 1996.[27] Social media platforms are not the authors of the content posted. They are online intermediaries connecting authors to viewers and allowing interaction. Platforms are more than infrastructure which connects people to the internet. While no social media platform operator would actively promote objectionable or extremist content, by design algorithms promote content that generates a lot of views to keep users engaged. Platforms like Facebook have never been similar to a phone company or a postal service.[28] Social media platforms are involved by design in the selection and dissemination of information.[29] They are not passive if their system distributes content.

---

[22] James Every-Palmer *Regulation of new technology: Institutions and processes* (2018, The Law Foundation New Zealand) at 11.

[23] Above at 12.

[24] Lynskey, above n 16, at 1.

[25] Above at 5.

[26] Above at 27.

[27] Andrew Murray "Rethinking Regulation for the Digital Environment" (2019) London School of Economics and Political Science, Policy Briefing 41 at 2.

[28] Cohen-Almagor, above n 21, at 435.

[29] Alexander Tsesis, "Social Media Accountability for Terrorist Propaganda" (2017) 86 FORDHAM L. REV. 605 at 623.

5 Platforms should not be left to self-regulate

Now internet platforms are thought of more like broadcasters than telecommunications operators due to the ability to curate content, but they are not regulated as broadcasters are. Platform operators have introduced self-regulation through content moderation to take some responsibility, but content moderation can be self-interested. Thompson quoted a 2017 report from the UK government which found that YouTube quickly removed material in breach of copyright but was slower to respond to otherwise illegal and hateful content.[30] The report stated in 2017, that the "biggest and richest social media companies are shamefully far from taking sufficient action to tackle illegal and dangerous content[31]". Every-Palmer describes an imbalance between technological change and regulatory change where internet companies and social media platforms are concerned.[32] There needs to be some external standards for platforms to follow otherwise, companies are left to make the rules about what content is unacceptable.

Self-regulation has its limitations and needs backing up with the real threat of legal sanctions that will hit the profit margins of platform operators if they do not take sufficient care. Censorship that is led by platforms rather than governments may be driven by economic motives. The Helen Clark Foundation has concerns about leaving platforms to regulate themselves as they are conflicted in that they profit from controversial content that creates high user engagement and greater advertising traffic.[33] This may be true for disinformation or discriminatory content designed to inflame political debate but objectionable content is more than controversial. There are reputational incentives for the large operators to minimise and remove objectionable content from their platforms by investing in content moderation staff and technology. It is in the interests of platforms to report and remove objectionable illegal content because a significant majority of users will find such content abhorrent and would not want to be using a platform that was seen to condone it. That is not to say that internet companies are devoid of morals as most people would not want to be associated with harmful content.

---

[30] Thompson, above n 2, at 85.

[31] House of Commons/Home Affairs Committee, 2017, at paras 3.1/3.3.

[32] James Every-Palmer "Regulation of new technology: Institutions and processes" (2018, The Law Foundation New Zealand) at 2.

[33] Mason and Errington, above n 3, at 9.

C Objectionable content

The worst types of content are deemed to be objectionable in New Zealand law. The FVPC Act is New Zealand's principal censorship legislation and provides a justified limitation on freedom of speech. The FVPC Act defines objectional publications as a publication that describes, depicts, expresses or otherwise deals with certain "gateway" subjects such as sex involving children, horror crime, cruelty or violence.[34] To be objectionable, making the publication available must likely be injurious to the public good and it must promote or support certain activities such as the exploitation of children for sexual purposes and acts of torture, extreme violence or extreme cruelty.[35] The main types of objectionable publication that must not be tolerated online and the focus of the regulation discussed in this paper are child sexual exploitation material and violent extremist content. Child sexual exploitation material has long been a focus of law enforcement in the online space and platform operators have responded to this repugnant content by deploying filters and algorithms to detect such material being posted on their platforms. Online violent extremism has not had the same recognition and it is arguably harder to define and detect due to the difference between legitimate journalism and terrorism promotion being context dependant.

Online violent extremism causes harm in many ways to viewers who are affected by the traumatic content and to victims and their families through re-victimisation. It can also lead to future acts of extreme violence through radicalisation when certain ideologies are promoted to susceptible people and can incite and teach how to carry out terrorist acts.[36] Online content influences offline behaviour, but the extent of this is complex and requires further research. Distinctions are often made between online and "in real life". With much more of life including business and socialising happening online, the online world is very much the real world. Terrorist or violent extremist content and child sexual exploitation material are without a doubt the two most problematic types of content online and are the easiest target for law enforcement. While other types of illegal content undoubtedly risk leading to physical world harm, taking wider government control of all content is much more difficult.

---

[34] Films, Videos and Publications Classifications Act 1993, section 3(1).

[35] Films, Videos and Publications Classifications Act 1993, section 3(2).

[36] Treasury *Regulatory Impact Statement - Films, Videos and Publications Classification (Urgent Interim Classification of Publications and Prevention of Online Harm) Amendment Bill 2020* <www.treasury.govt.nz>.

III New Zealand's lead up to change

New Zealand is not immune to the online risks that have emerged overseas. Prior to 15 March 2019, there was a naivety that New Zealand was isolated from global problems. Research from the Department of Internal Affairs shows that New Zealanders are consuming a "full spectrum" of what can be described as "extremist content" ranging from far left-wing to far right-wing and including Islamic extremism, environmental extremism and conspiracy theories. However, per capita the problem in New Zealand is no worse than overseas.[37] John Edwards commented that the New Zealand government swiftly moved to ban semi-automatic weapons following March 15 2019, however there was no corresponding effort to ban social media, a key tool used by the terrorist in disseminating the attack.[38] New Zealand has taken a slow and considered approach to changing censorship laws towards the introduction of take-down powers and removing the ability for platforms to claim safe harbour from liability. The following chapter discusses the surrounding work leading to the proposed changes to the FVPC Act.

A Royal Commission report

The report of the Royal Commission of Inquiry into the terrorist attack on Christchurch mosques on 15 March 2019 *Ko tō tātou kāinga tēnei* was presented to Parliament on 8 December 2020.[39] This paper only considers the parts of the report concerned with the online anti-terrorism capabilities of New Zealand government agencies and New Zealand's censorship laws. The Royal Commission did not deeply analyse the role of social media platforms other than the way they were used by the terrorist. The Royal Commission noted that 15 March highlighted the role of online counter-terrorism responsibilities of government agencies at the internet is a key platform not only for radicalisation, recruitment and funding but also dissemination of terrorist acts such as via livestreaming.[40] Significant challenges were identified in online monitoring and enforcement due to the size and complexity of the internet, increased use of encryption, anonymity, rapid changes in technology and the boundaries between free speech and harmful behaviour.[41]

---

[37] Laura Walters "Govt research: NZ importing right-wing extremist content" (Newsroom, 13 January 2021) <www.newsroom.co.nz>.

[38] John Edwards, above at 19.

[39] Royal Commission of Inquiry into the Attack on Christchurch Mosques on 15 March 2019 *Ko tō tātou kāinga tēnei Report: Royal Commission of Inquiry into the terrorist attack on Christchurch masjidain on 15 March 2019* (8 December 2020).

[40] Above at part 8, chapter 11 at [2].

[41] Above at part 8, chapter 11 at [9].

Prior to 15 March 2019, New Zealand government agencies had limited monitoring and enforcement capability for online counter-terrorism. The Royal Commission heard evidence from the New Zealand Police that there was very little training and few tools to assist the Police in social media intelligence collection. Since March 2019, the New Zealand Police now has a dedicated team and tools for rapid data extraction.[42] The Department of Internal Affairs, the Government Communications Security Bureau, New Zealand Police and the New Zealand Security Intelligence Service) all have a role in relation to extremist activity online. These agencies have different mandates, while enforcement agencies may seek to shut down objectionable content, counter-terrorism agencies, counter-terrorism agencies may seek to monitor accounts for a prolonged period to collect intelligence on a person's intent and capabilities.[43] The Royal Commission expressed concern that capabilities are expanding without a whole of system perspective and called for greater oversight and coordination of resourcing and objectives.[44] As capability grows, it will be important to avoid duplication of work, use resources efficiently and avoid conflicting objectives by developing a clear shared understanding of the legal and policy settings and social licence.[45]

In 2019 it was announced that Department of Internal Affairs would receive a funding boost of $17 million over four years to increase the capacity of its Censorship Compliance Unit which had previously been resourced to combat online child exploitation.[46] Addressing online extremism largely relies on the voluntary cooperation of platform operators. Currently, law enforcement relies on platforms removing illegal content on request as the platform operators have no liability as content hosts. Prior to 15 March 2019, objectionable publications promoting or supporting terrorism were not a primary focus of agencies enforcing the FVPC Act due to funding and mandate constraints.

B Christchurch Call

On 15 May 2019, New Zealand Prime Minister Jacinda Arden and French President Emmanual Macron convened a summit in Paris with 17 state government representatives and eight leaders from major technology companies which would become the Christchurch

[42] Above at part 8, chapter 11 at [21] and [26].

[43] Above at part 8, chapter 11 at [30].

[44] Above at part 8, chapter 11 at [32]-[34].

[45] Above at part 8, chapter 11 at [37].

[46] Marc Daalder "Government to boost efforts against online extremism" (Newsroom, 14 October 2019) <www.newsroom.co.nz>.

Call.[47] The Christchurch Call is now supported by 38 other states. Notably the US was initially absent, however, in May 2021 the US announced that it was joining the Christchurch Call. In a statement, the US deemed that countering online radicalisation and recruitment by terrorists and violent extremists was a significant priority while at the same time needing to protect freedom of expression.[48]

The Christchurch Call is a list of voluntary commitments for governments and technology companies to address terrorist and violent extremist content online without compromising a free and open internet, human rights and freedom of expression.[49] The Technology companies have given undertakings to use technology to prevent terrorist and violent extremist content from being uploaded and to enforce their terms of service responsibly.[50] States have made commitments to enforce laws prohibiting such material while respecting the rule of law and human rights.[51] Thompson is sceptical about the motivations for platforms expressing willingness to be subjected to state regulation as the areas of agreement between governments and companies are narrow.[52] In his view, the concerns about extremism while important to address, are symptoms of structural issues in the digital ecology linked to the financial models of platforms. Ip, however, describes it as a "small but discernible shift in the landscape of digital responsibility[53]". While the Christchurch Call has brought governments and companies together, it is only voluntary commitments and states must follow through with making domestic laws to hold companies to account.[54]

The Christchurch Call has been criticised for not doing enough to address the causes of extremist content. It is true that the Christchurch Call does not directly target the source of the problem, but it is an attempt to reign in technology companies which have inadvertently amplified terrorism and violent extremism. The causes are a lot more complicated than what can be achieved in one summit. If the Christchurch Call is treated as a starting point, there is potential to expand the work. Anjum Rahman, one of the co-chairs of the Christchurch Call Advisory Network, believes that to be meaningful, the Christchurch Call

[47] Christchurch Call <www.christchurchcall.com>.

[48] Statement by Press Secretary Jen Psaki on the Occasion of the United States Joining the Christchurch Call to Action to Eliminate Terrorist and Violent Extremist Content Online <www.whitehouse.gov>

[49] Above

[50] John Ip "Law's response to New Zealand's 'darkest of days'" (2021) 50 Common Law World Review 1 21–37 at 27.

[51] Above at 28.

[52] Thompson, above n 2, at 91.

[53] Ip, above n 50, at 28.

[54] Thompson, above n 2, at 92.

must go beyond its current consensus that livestreaming mass murder is wrong.[55] In Rahman's view, the work of the Christchurch Call needs to extend to addressing the online influences and groups where the sharing of hate speech content precedes terrorist acts and hate crimes.[56] Thompson also questions the level of appropriate regulatory action, the Christchurch Call is focussed on preventing online dissemination of terrorist content while the ultimate goal should be reducing terrorist acts.[57] The Helen Clark Foundation has produced reports on this topic (discussed later in this paper) and supports the Christchurch Call but believes that social media requires coordinated and comprehensive regulation.[58]

Whether social discord is the root of online extremism or not, it is amplified by the profit models and algorithmic designs of social media platforms to disseminate content that will gain attention regardless of its morality. While the Christchurch Call is only addressing the symptoms of deeper issues, it represents a step towards changing the power dynamic between online platforms and states. If the dissemination of online terrorist and violent extremist content is likely to inspire other would-be terrorists then preventing dissemination should go some way to reducing physical acts and reducing trauma for victims.

IV Proposed changes to the FVPC Act

The Films, Videos and Publications Classification (Urgent Interim Classification of Publications and Prevention of Online Harm) Amendment Bill 2020 (the Bill) was introduced in May 2020. The intention of the Bill is to fill in gaps identified by the Royal Commission and support the Christchurch Call commitments to give law enforcement the necessary powers to rapid respond. The Bill will:

> ensure effective enforcement of applicable laws that prohibit the production or dissemination of terrorist and violent extremist content, in a manner consistent with the rule of law and international human rights law, including freedom of expression.[59]

The Bill will go some way to addressing the current void of responsibility for the online content hosts whose technology allows for the dissemination of objectionable publications. The Regulatory Impact Statement noted that as implementing changes to respond to 15

[55] Anjum Rahman "A year on, the Christchurch Call must go beyond 'don't livestream mass murder'" (The Spinoff, 15 May 2020) <www.thespinoff.co.nz>.

[56] Above.

[57] Thompson, above n 2, at 94.

[58] Mason and Errington, above n 3, at 5.

[59] Regulatory Impact Statement, above n 36.

March 2019 is a government priority, the policy development for the Bill has been under significant time pressure.[60] This is not urgent legislation, with the Bill likely to pass in late 2021, about a year a half after the Christchurch Call. It has been met with some opposition in submissions mainly due to the proposal to include the ability to create mandatory internet filters through regulations. Internet NZ summed up the concerns in their submission, urging caution against blocking at ISP level which they believe will not work and will likely have unintended consequences in blocking of legitimate content, threating open access to the internet and causing human rights issues.[61] The Select Committee report recommended removing all references to electronic filters.[62] The main changes remaining in the Bill are enabling interim classification decisions, creating an offence of livestreaming an objectionable publication, removing safe harbour for online platforms and empowering authorities to issue take-down notices.

A Classification powers

The powers of the Chief Censor will increase to enable the Classification Office to make interim classification assessments to quickly notify the public that specific content is likely to be objectionable when there is an urgent need prior to issuing a full written decision.[63] Interim classification decisions will last for 20 days. There will be immunity from civil and criminal liability for officials in issuing an interim classification and for online content hosts in relying on an interim classification to remove content.[64] Interim classifications are one potential basis for take-down notices.

B Livestreaming

Livestreaming an objectionable publication will also become a specific offence in addition to the current relevant offences for making and distributing an objectionable publication.[65] A definition of 'livestream' is also included in the Bill to capture images or sounds transmitted or streamed over the internet as they happen. This is an example of a response to new technology which was not clearly captured under traditional forms of publishing

[60] Above.

[61] InternetNZ "To block or not to block Technical and policy considerations of Internet filtering" (September 2019) <www.internetnz.nz>.

[62] Films, Videos, and Publications Classification (Urgent Interim Classification of Publications and Prevention of Online Harm) Amendment Bill 2020 (268-2) (select committee report) at 3.

[63] Films, Videos and Publications Classification (Urgent Interim Classification of Publications and Prevention of Online Harm) Amendment Bill 2020 (268-2), clause 6, new section 22A.

[64] Clause 6, new sections 22C and 22D.

[65] Films, Videos and Publications Classification (Urgent Interim Classification of Publications and Prevention of Online Harm) Amendment Bill 2020 (268-2), clause 10.

content. This amendment is intended to be targeted at people who livestream objectionable content. The Select Committee recommended extending it to those who share the link to the livestream.[66] Once an objectionable livestream has occurred it becomes an objectionable publication with is illegal to possess or distribute so it is not clear why this is necessary. The Select Committee rightly notes that intent is difficult to determine while content is being livestreamed and there is risk of legitimate news reporting being captured. The amendment includes knowledge of likelihood of objectionability and the intent of promoting or encouraging criminal acts or acts of terrorism as elements of the offence.[67] The addition of livestreaming in the Act has a risk of creating a precedent for a need for the legislature to respond to every development in technology that was not clearly within the definition of publication. This could have future repercussions, affecting what types of publications can be subject to take-down notices.

C No safe harbour

One of the objectives of the Bill is to ensure that the 'safe harbour' provisions in the Harmful Digital Communications Act 2015 (HDCA) do not apply to the FVCPA to bring social media platforms within scope.[68] The Bill will not amend the HDCA, it only clarifies that the new Part 7A relating to take-down notices will be exempt from the safe harbour process.

The HDCA applies to communications where an individual victim in targeted and harm is caused. While it encompasses a broad range of digital communications, it does not apply to communications which denigrate groups and where harm cannot be made out.[69] The purpose of the Act is directly focussed on individual victims.[70] Generally, online content hosts can rely on the protection from liability for user generated content referred to as 'safe harbour' if they follow the process in section 24 when they receive a complaint.[71] The

---

[66] Films, Videos, and Publications Classification (Urgent Interim Classification of Publications and Prevention of Online Harm) Amendment Bill 2020 (268-2) (select committee report), at 7.

[67] Films, Videos and Publications Classification (Urgent Interim Classification of Publications and Prevention of Online Harm) Amendment Bill 2020 (268-2), clause 10, new section 132C.

[68] Films, Videos and Publications Classification (Urgent Interim Classification of Publications and Prevention of Online Harm) Amendment Bill 2020 Bills Digest 2626 (15 June 2020);
Films, Videos and Publications Classification (Urgent Interim Classification of Publications and Prevention of Online Harm) Amendment Bill 2020 (268-2), clause 5.

[69] Royal Commission of Inquiry into the Terrorist Attack on Christchurch Mosques on 15 March 2019 *Hate speech and hate crime related legislation* (26 November 2020).

[70] Harmful Digital Communications Act 2015, section 3.

[71] Above, sections 23-25.

process gives the author of potential harmful or illegal content a right of response within 48 hours. If no response is received, the online content host must remove it, but if the author disagrees with removal, it is up to the host to decide whether it breaches its terms and conditions warranting removal. Under this model, online content hosts are not making a judgment whether the content is harmful or illegal. Safe harbour extends to any civil or criminal proceedings beyond the HDCA so online content hosts can rely on it if they follow the same process regarding objectionable publications are posted on their platforms so they will not be liable under the current provisions in the FVCPA. There is also some potential for action to be taken against platforms under the HDCA. The District Court can make orders against online content hosts[72] for example to take down or disable access to content, to have the author of content posted anonymously or under a pseudonym identified.[73]

D Take-down notices

The Bill proposes that Inspectors of Publications may issue take-down notices to online content hosts for particular publications if the publications are subject to an interim objectionable classification, classified as objectionable or believed by the Inspector on reasonable grounds to be objectionable.[74] The Inspector can decide what time period is reasonable to take down the content. This effectively gives Inspectors censorship powers. The new take-down powers in the FVPCA will not require a court order like the take down powers under the HDCA and the Contempt of Court Act 2019. Take-down powers are aligned with current powers for the seizure of objectionable publications in the FVPCA.[75] The Classifications Office currently can accept submissions from certain law enforcement officials and from any others with leave of the Chief Censor.[76] The amendment will also allow online content hosts who are subject to a take-down notice to submit online publications for a classification determination.[77] A content host could seek a classification decision where an Inspector has determined something likely to be objectionable and then have the decision reviewed again if the content host does not

[72] online content host is defined in section 4 of the HDCA - in relation to a digital communication, means the person who has control over the part of the electronic retrieval system, such as a website or an online application, on which the communication is posted and accessible by the user.

[73] Harmful Digital Communications Act 2015, section 19(2).

[74] Films, Videos and Publications Classification (Urgent Interim Classification of Publications and Prevention of Online Harm) Amendment Bill 2020 (268-2), clause 9, new section 119C.

[75] Films, Videos, and Publications Classification (Urgent Interim Classification of Publications and Prevention of Online Harm) Amendment Bill (departmental disclosure).

[76] Films, Videos and Publications Classifications Act 1993, s 13.

[77] Films, Videos and Publications Classification (Urgent Interim Classification of Publications and Prevention of Online Harm) Amendment Bill 2020 (268-2), clause 5A, new section 13(1)(ba).

agree. The existing review process in the Part 4 of the FVPCA will also apply to the classification decision that relates to the take-down notice.[78] The host can preserve the content pending the outcome of a review of the classification. If an online content host does not seek review or comply, proceedings may be taken in the District Court, however the Court must not examine the merits of the notice and may only consider whether the content host had a reasonable justification for failing or refusing to comply. The Court may make a number of orders including for compliance with the take-down notice by a certain date and can issue a pecuniary penalty up to $200,000. The new Part 7A will apply to online content hosts in New Zealand and overseas who provide services to the public.

In practice, it is likely that existing practice of voluntary requests to take down objectionable content would be relied upon primarily, particularly when operators are based overseas.[79] The Bill acknowledges that Inspectors may, but are not required to, request that the content be removed or have access prevented before issuing the notice. The Department of Internal Affairs has indicated an intention to use the take-down process where other options for having content removed are ineffective.[80] The Departmental Disclosure for the Bill indicates that existing relationships with platforms will be maintained and that the take-down notices are merely a formalisation of existing practices.

The proposed maximum penalty of $200,000 is higher than the available financial penalties in the criminal jurisdiction however remarkably low considering the power and revenue of large social media operators. Despite the good intentions of online content hosts to address objectionable content on their platforms, some form of enforceable penalty is preferrable to take content hosts out of the realm of self-regulation.

There are risks associated with the proposed legislation. The first is that platforms may become overly restrictive and remove content that is not objectionable rather than wait for take-down notices. Take-down notices may be unnecessary or ineffective if large platforms already have their own systems in place for the reporting and removal of content.[81] Where content is widely disseminated rapidly and in different edited versions making complete

[78] Films, Videos and Publications Classification (Urgent Interim Classification of Publications and Prevention of Online Harm) Amendment Bill 2020 (268-2), clause 9, new section 119J.

[79] Marc Daalder "New bill comes with online takedown powers" (Newsroom, 27 May 2020) <www.newsroom.co.nz>.

[80] Departmental disclosure, above n 75.

[81] Regulatory Impact Statement, above n 36.

removal technically impossible, a take-down notice will not expediate removal. Take-down notices must contain a description of the relevant publication and a URL or other unique identifier.[82] Requiring the level of specificity of a URL or identifier means that notices must be issued for individual publications rather than anything depicting a particular event. This makes compliance easier for platforms because if a notice was issued for every depiction of the terror attack of 15 March 2019, compliance would be a constant task as new versions continue to appear online. Ideally, platforms will use a take-down notice as a license to remove any copies or edited versions of the same content appearing on other URLs or with different fingerprints or hashes, otherwise the burden on the regulator is too high to respond to viral content. The capacity to comply with a take-down order that references an identifier unique to the file rather than its location would require some form of algorithm or automated tool. Smaller platforms may choose to ignore any notices issued and wait to be formally pursued by court order, leaving harmful content available. Extremists may be given further reason to move to smaller and less cooperative platforms or to move more to peer-to-peer file sharing platforms[83] which cannot be subject to the take-down notice and civil penalty processes.

V Enforcement against overseas based internet companies

A Making extraterritorial law

The proposed amendments to the FVPC Act in the Urgent Classification Bill will apply to "online content hosts both in New Zealand and overseas that provide services to the public." Take-down orders will have extraterritorial effect. While it is possible to make extraterritorial laws, whether enforcement is practicable is another question. New Zealand's parliament has the full power to make laws including outside of New Zealand but there is common law presumption that the law does not apply outside New Zealand unless it explicitly says so.[84] Traditional territorial jurisdiction needs to be expanded to regulate the online environment. When users in New Zealand are at risk of harm from viewing content, there is a jurisdictional link to justify intervention.

There must be a clear case for New Zealand law to apply, and it must be reasonable to expect the people to whom the legislation will apply to comply with New Zealand law. The

[82] Films, Videos and Publications Classification (Urgent Interim Classification of Publications and Prevention of Online Harm) Amendment Bill 2020 (268-2), clause 9.

[83] Encrypted peer-to-peer online messaging services which allow forming groups and channels to disseminate information are harder to detect as law enforcement needs to obtain devices with passcodes.

[84] *Poynter v Commerce Commission* SC 32/2009 [2010] NZSC 38 at [15].

Legislation Design and Advisory Committee (LDAC) note that any proposal to apply criminal law extraterritorially is relatively unusual and subject to unique considerations. Offences should only have extraterritorial application in exceptional circumstances.[85] LDAC's view is that offences are one of a variety of tools and should only be created where there are no reasonable alternatives such as self-regulation by the applicable industry, non-criminal state measures, such as education campaigns, informal warnings, or other methods of persuasion, such as codes of practice or national standards; or other forms of State enforcement, such as civil remedies including pecuniary penalties.

Some New Zealand legislation specifies extraterritorial effect. The Privacy Act 2020 applies to any action taken by overseas agencies in the course of carrying on business in New Zealand in respect of personal information collected or held by the overseas agency. The definition of "carry on business" is broad. Section 4(3) makes it clear that an agency may be treated as carrying on business in New Zealand without necessarily having a physical place of business in New Zealand. The extraterritorial application for Commercial Video on Demand (CVoD) providers in the FVCP Act is:[86]

> in respect of commercial video on-demand content that is made available in New Zealand by a specified CVoD provider regardless of whether the provider is resident or incorporated in New Zealand or outside New Zealand.

The definition of carrying on business in New Zealand in the Companies Act 1993 is different and is for the purpose of a requirement to have a registered office.[87] Online platforms do not need to registered in New Zealand to offer their services to New Zealanders.

B Pecuniary penalties

Extraterritorial civil penalty regimes are generally chosen by the legislature as more appropriate to regulate overseas corporate activity which affects New Zealanders. The proposed online pecuniary penalty in the Bill is $200,000 for failure of an online content host to comply with a take-down order in the new section 119I. A pecuniary penalty is a punitive measure designed to punish and deter contravention. Penalties can be imposed following a trial under the rules of civil procedure and evidence. Liability is to the civil standard of proof, on the balance of probabilities. The penalty can only be monetary, and

[85] Legislation Design and Advisory Committee *Legislation Guidelines* (2018) <www.ldac.govt.nz> at chapter 24.

[86] Films, Videos and Publications Classifications Act 1993, s 46C.

[87] Companies Act 1993, section 332.

it is paid to the Crown. Pecuniary penalties are appropriate for strict liability regulatory offences which are commercial in nature and are more appropriate than the criminal jurisdiction where there may be complex issues of proof.

The maximum penalty for non-compliance with a take-down order is low considering the size and financial capacity of the likes of Facebook and Google. Compliance becomes essentially voluntary if penalties and debt recovery are difficult and resource intensive to enforce. The incentive to comply for large platforms may be largely reputational. It is difficult to comment on the appropriate maximum penalties for civil liability acts in by social media platforms. There is precedent in New Zealand legislation for penalties into the millions of dollars.[88] The LDAC Guidelines provide some guidance on setting penalties.[89] The maximum penalty should not be disproportionately severe, but should reflect the worst case of possible offending. The penalty for non-compliance with a take-down notice does not seem to reflect the worst-case scenario of a blatant refusal to remove objectionable content.

C Service

For a civil claim to proceed, the proceedings need to be served. If the company responsible does not have a physical presence in New Zealand, then service will need to take place overseas. Service on an overseas defendant can take place under the High Court Rules 2016 with or without leave of the court.[90] For laws applying to social media platforms, it needs to not matter where the offence takes place or where the company is based.

Using Facebook as an example, Facebook has a registered company called Facebook New Zealand Limited that is a subsidiary of US based Facebook, Inc. with an address for service as a Wellington based law firm and directors based in Australia, Ireland and Singapore.[91] When asked by a journalist, Facebook would not comment on the size of focus of the New Zealand office and former employees indicated that the focus was on advertising.[92] While it would be easier to file proceedings against local subsidiaries, they are not likely

[88] For example, the Anti-Money Laundering and Countering Financing of Terrorism Act 2009 has penalties up to $2 million for civil liability acts and fines up to $5 million for criminal offences, see sections 90 and 100.

[89] LDAC Guidelines, above n 85, at chapter 26.

[90] High Court Rules 2016, Part 6 subpart 12.

[91] New Zealand Companies Office, Companies Register FACEBOOK NEW ZEALAND LIMITED *(3043328) Registered* <app.companiesoffice.govt.nz>.

[92] Henry Cooke "The disparate global locations that make up New Zealand Facebook - and your news feed" (Stuff, 27 March 2017) <www.stuff.co.nz>.

responsible for content moderation or setting company policy and will not have significant assets.[93] Enforcement proceedings would likely need to be filed against and served on the US based parent company. It seems to be a trend that legislation that applies to online activity by overseas companies which affects New Zealanders does not require those companies to have a New Zealand registered office in order to assert jurisdiction.

D Enforceability

Extraterritorial offences often rely on international agreements and cooperation between states. The internet does not have borders and platforms spread their operations across physical jurisdictions. It is noted in the Bill's Regulatory Impact Statement that in practice, a penalty for failure to comply with a take-down notice may not be enforceable outside of New Zealand and platform operators may view themselves as immune from liability.[94] If there is no intention to enforce the law, then it is difficult to see what the point of making it is. The internet platforms the law is targeted at are based overseas. There is legislation for reciprocal enforcement of judgements with Australia, the Reciprocal Enforcement of Orders Act provides for the enforcement of New Zealand civil orders in Australia and the Trans-Tasman Legal Proceedings Act 2010. However, this would generally not assist with the locations of large internet platform operators.

Reputational risk is likely to be a significant incentive for platforms to comply with New Zealand law. The Regulatory Impact Statement notes that even if the new laws are extremely difficult to enforce, the New Zealand government is still sending a strong message that platforms should take responsibility for content. Implementation is reliant on ISPs and social media platforms cooperating as "good corporate citizens", political and public support and adequate operational resources to use the new powers.[95] The Select Committee report also considered extraterritoriality and enforceability and noted that New Zealand would need formal agreements with other countries for enforcement, however take-down notices would make cooperation more likely.[96] The heavy reliance on cooperation from platforms is of concern as this willingness to cooperate could fade over time or be outright reversed. If there is a serious intention to hold platforms to account,

---

[93] Jasmine Valcic "The Sharing of Abhorrent Violent Material Act: The Realities and Implications of Australia's New Laws Regulating Social Media Companies" (2021) 33 Bond Law Review 1 11-35 at 26.

[94] Regulatory Impact Statement, above n 36.

[95] Regulatory Impact Statement, above n 36.

[96] Films, Videos, and Publications Classification (Urgent Interim Classification of Publications and Prevention of Online Harm) Amendment Bill 2020 (268-2) (select committee report) at 2.

then resources must be put into issuing take-down notices and taking enforcement action where necessary.

E Budapest Convention

The Council of Europe Convention on Cybercrime (the Budapest Convention) came into force in 2004 and is the first treaty seeking to address internet and computer crime. It provides a framework for global cooperation to align domestic legislation to enable the sharing of intelligence and electronic evidence for investigations. On 2 June 2020, Cabinet agreed to New Zealand's accession to the Budapest Convention pending further public consultation.[97] New Zealand's laws are mostly aligned with the Convention and only minor amendments are necessary to improve reciprocal assistance to partners.[98] The Convention also requires its implementation to be consistent with human rights, freedom of expression and privacy protections.[99] Cooperation extends to any crime facilitated through technology including online child exploitation and online violence extremism in addition to more common cybercrimes such as fraud and phishing scams. Electronic evidence can be found through social media communications and cloud storage data.[100] A key feature is the need for the ability to serve data preservation orders on platforms so that content can be preserved as evidence to be used in investigations into users and the ability to apply for surveillance device warrants on behalf of overseas agencies.[101] Accession ensures that New Zealand can be well informed of global development in countering online crime and has the best chance of enforcing domestic laws against online platforms. The Select Committee noted that the Convention could enable a formal avenue for enforcing take-down notices in signatory countries.[102] This would be a welcome addition although the Convention is mostly focussed on evidence collection and investigation.

---

[97] Department of the Prime Minister and Cabinet and Ministry of Justice *Cabinet Paper: Budapest Convention on Cybercrime: Approval to Initiate the First Stage Towards Accession* (2020).

[98] Above at [15].

[99] Above at [18].

[100] Above at [20].

[101] Above at [34].

[102] Films, Videos, and Publications Classification (Urgent Interim Classification of Publications and Prevention of Online Harm) Amendment Bill 2020 (268-2) (select committee report) at 3.

F Commerce Commission v Viagogo AG[103]

The ongoing litigation taken by New Zealand's Commerce Commission against Viagogo highlights the difficulty in extraterritorial law enforcement. Viagogo is an online ticket resale platform based in Switzerland. In November 2018 the Commerce Commission claimed that Viagogo was making false, misleading or deceptive representations to New Zealand consumers and applied for an interim injunction to stop Viagogo from continuing to operate in breach of the Fair Trading Act 1986. Viagogo declined to accept service of the proceedings through its New Zealand Lawyers and required proceedings to be served at its headquarters in Switzerland. Viagogo then told the Commission that if proceedings were served in Switzerland, it would challenge the jurisdiction of the New Zealand courts. The Commission proceeded to arrange service of the proceedings on Viagogo in Switzerland through diplomatic channels. The court noted that this process can take many months.[104] The Court of Appeal decided that an application for injunctive relief could be heard without notice before service of the substantial proceedings. When the matter went to the Court of Appeal, Viagogo had accepted service but challenged the jurisdiction.[105]

This decision did not address whether Viagogo was engaging in conduct in New Zealand or carrying on business in New Zealand for the purposes of the Fair Trading Act. Section 3 of the Fair Trading Act provides that the Act "extends to the engaging in conduct outside New Zealand by any person resident or carrying on business in New Zealand to the extent that such conduct relates to the supply of goods or services, or the granting of interests in land, within New Zealand." This is still at issue even though Viagogo has now submitted to the jurisdiction of the High Court. This case does demonstrate that New Zealand authorities can pursue overseas entities if they have the will and the resources.

VI What else can be done to regulate content providers?

There is potential to do more than just take-down orders to place the burden on compliance on platforms who offer their services in New Zealand. The different approaches which can overlap or be used in combination are discussed in this section. The level of responsibility to place on platforms must be balanced with what is practicable to comply with. Scholars have examined different regulation models ranging from industry self-regulation to mixed government and private models and full government control. Much of the literature

[103] *Commerce Commission v Viagogo AG* [2019] NZCA 472.

[104] Above at [2]

[105] Above at [9]

discussed below is based on hypothetical risks. Internationally, laws compelling platforms to monitor or remove content are in the early stages of implementation. As the Workshop's research found, there are many ideas about how to regulate platforms, but no tested solutions backed up by long term data.[106] If widely regulating platforms is urgent, there must be some element of boldness and experimentation in the law.

A Delegation of responsibility to platforms

Governments can delegate responsibility to platform operators by requiring them to carry out censorship functions rather than relying on a monitoring role and using take-down powers. Governments can use the resources of platforms to administer and enforce laws that would otherwise require significant government investment. The large volume of content requiring moderation and the technology and staffing needed mean it is practical to put the responsibility onto platforms. While cost effective, delegating power or requiring self-regulation creates a risk of internet platforms slipping into a blind spot if there is no transparency.

Delegation to intermediary platforms is "highly problematic" due to the loss of accountability and transparency.[107] Land has considered the risks of "privatised censorship" from an international human rights law perspective.[108] The decisions made by platforms operators have significant effects and can lead to human rights concerns such as freedom of expression. Errington commented that the public does not get a say in who the CEO of and board of directors of Facebook and Google are and there is no public scrutiny or consultation involved when they make rules.[109] A delegated censorship or duty of care model could require compulsory deployment of automated censorship tools. National security and protecting the public is traditionally the role of the state. Macdonald et al. argue that platforms should comply with human rights and rule of law principles as well as moral and social responsibilities.[110]

---

[106] The Workshop, above n 14, at 17.

[107] Molly K. Land "Against Privatized Censorship: Proposals for Responsible Delegation" (2020) 60:2 Virginia Journal of International Law 363, at 389.

[108] Above at 365.

[109] Katherine Errington, in conversation with Anjum Rahman "Reducing online harm" in Andrew Chen (ed.) Shouting Zeros and Ones: Digital Technology, Ethics and Policy in New Zealand (2020, Bridget Williams Books, online ed.) at [41].

[110] Stuart Macdonald, Sara Giro Correia and Amy-Louise Watkin "Regulating terrorist content on social media: automation and the rule of law" (2019) International Journal of Law in Context 15, 183–197 at 186.

It is important that platforms are reporting content that is illegal or suspected to be illegal to the relevant authorities and preserving the evidence. Removing the content without further consequences does not address the root of the problem and leaves victims without redress. While it is practical from a resourcing perspective to delegate censorship power, there still needs to be a right of appeal of removal and compulsory linking with government authorities where serious criminal activity is detected. A safe harbour for platforms is still possible within a delegated responsibility framework if platform operators can demonstrate that they took steps to remove content within a reasonable or specified time period.[111]

A privatised censorship role could be useful for content that clearly falls within the definition of objectionable. Land proposes a form of "differentiated liability" as a solution. Censorship power could only be delegated for only the most serious types of illegal content where there is high likelihood of harm and low risk to reasonable freedoms of expression.[112] Differentiated liability combined with government oversight and user-centric design would provide a balanced and realistic regulatory approach.[113] Social media platforms such as Facebook are already removing objectionable content without any right for users to request a review by a government agency. A mixed model of state regulation and self-regulation by companies is desirable because it is impractical for the state to be the sole monitor of social media and there are risks in requiring social media companies to be the sole deciders of what content stays.

B Duty of care

A duty of care model would be more prescriptive than a delegation of censorship responsibilities. A primary duty of care model is used in New Zealand's Health and Safety at Work Act 2015. Persons in control of a business or undertaking are required to so far as is reasonably practicable ensure the safety of workers and must ensure, so far as is reasonably practicable, that the health and safety of other persons is not put at risk from the work.[114] The second part of the primary duty of care relating to others could be a model for a social media duty regarding harm that is likely to come about from tools or business models such as livestreaming and targeted content. Platforms could be required to so far as is reasonably practicable ensure that their business activities do not harm users. A

[111] Land, above n 105, at 381.

[112] Above at 421.

[113] Above at 421.

[114] Health and Safety at Work Act 2015, s 36.

government regulator could monitor platforms by requiring reporting with intervention powers where necessary.

The health and safety at work model is not directly analogous to social media as the groups of people requiring protection is all users. Users have diverse needs and expectations in how they interact with social media. It does have relevance from the angle of the needs of business to be balanced with the safety of individuals. A duty of care model would require platform operators to take ownership of responsibility for content and invest in systems or personal to meet compliance standards. This requires assessment of the risks and the reasonably practicable steps to eliminate or minimise them. Edwards favours some form of duty of care because other industries products must be safety tested for safety prior to being used. The potential for misuse of livestreaming technology was not well thought through.[115] Arguably there should be a duty of care to take all reasonable steps to prevent, reduce or mitigate the harmful effects of products.[116]

The purported benefits of imposing a duty of care are that rather than focussing on the removal of specific content, the law is focussed on a goal of keeping users safe by imposing process requirements on the platform operator.[117] The UK Online Harms White Paper[118] was published in 2019 and will form the basis of a new regulatory framework, the Online Safety Bill 2021. The White Paper recommends duty of care responsibilities for companies to take responsibility for the safety of users following codes of practice set by a regulator with transparency, accountability and trust being critical elements.[119] The companies in scope are those who facilitate the sharing and discovering of user generated content and allow users to interact.[120] The recommended responsibilities include having an easy-to-access complaint function, independent review processes and specific obligations to take steps to combat the most harmful content (notably extremist/terrorist material and child sexual abuse).[121] This model would require significant investment by social media providers in their content moderation technology. The White Paper does not go as far as recommended processes such as take-downs by prescribed in legislation, it favours the

[115] Mason and Errington (Q and A with John Edwards), above n 3, at 23.

[116] Edwards, above at 19.

[117] Lorna Woods "The duty of care in the Online Harms White Paper" (2019) 11 Journal of Media Law 1, 6-17 at 11.

[118] Department for Digital, Culture, Media and Sport, and the Home Office *Online Harms White Paper* (2019).

[119] Above at 41.

[120] Above at 49.

[121] Above at 8.

details being in codes.[122] A broad duty allows for flexibility for secondary guidance to be developed for specific types of content. Murray, a lecturer at the London School of Economics, describes the duty of care model as "lurching from under-regulation to over-regulation[123]" due to the overwhelming scope. While there is good intentions behind it, a duty of care is risky as online content is broad and user controlled. It is not analogous to narrow physical industries that can take care of safety in their areas of influence and control.

The role as gatekeeper comes into duty of care discussions, as it is the platforms who effectively decide what is in or out. While governments make certain content illegal, the compliance or cooperation of platforms is key to effective control. In practice, platforms are largely responsible for deciding what content stays and goes and which users are allowed to participate.[124] Without sufficient oversight, a duty of care could be privatised censorship by another name. For a duty of care to be effective, there needs to be industry best practice guidance from governments or international bodies so that platforms apply consistent definitions of terrorism and violent extremism which is harder to define than child exploitation.[125]

C An independent regulator

New Zealand does not have a single regulator for online platforms. Responsibility is spread over multiple agencies. Reports in the UK and France have recommended independent regulators. Mason and Errington are concerned about the "piecemeal fashion" of different aspects of social media being regulated by multiple parts of government including the Privacy Commission, the Ministry of Justice, the Department of Internal Affairs, and Netsafe.[126] Establishing a dedicated agency to bring together online regulatory is an attractive idea but could have an unworkable scope.

An independent regulator could be tasked with setting rules and standards and taking enforcement action for non-compliance. The Helen Clark Foundation recommends that New Zealand establish an independent regulator to oversee social media companies similar to the Broadcasting Standards Authority or the New Zealand Media Council.[127] This could be either an independent Crown Entity, the industry could be required to set up its own

[122] Woods, above n 115, at 16.

[123] Murray, above n 27, at 5.

[124] Lynskey, above n 16, at 13.

[125] Roter, above n 11, at 1408.

[126] Mason and Errington, above n 3, at 4.

[127] Above at 16.

body without enforcement powers or a combined model where an industry body could issue codes of practice and recommendations.[128] The work of an independent regulator could be funded through cost recovery levies as a sort of licence to offer online content services to the public in New Zealand. The problem with this as a proposal for New Zealand is that the current "patchwork" of social media regulation would not easily lend itself to collecting levies and creating an overall compliance landscape. Another idea Every-Palmer proposes is a centralised government function for adaptation to new technology which looks for emerging issues caused by technology and coordinates the regulatory response.129 A social media regulator or general internet regulator could have this role and anticipate potential problems with new technology such as livestreaming before an incident occurs.

The UK Online Harms White Paper also recommends an independent regulator to implement, oversee and enforce compliance with the proposed duty of care.[130] The UK has a Counter-Terrorism Internet Referral Unit (CTIRU) which employs staff to review extremist content on Facebook that breaches terms of service and then asks Facebook to remove it.[131] The US hosts the National Centre for Missing and Exploited Children (NCMEC), which US based social media platforms report child exploitation material to and NCMEC passes on ISP information to the relevant international authority.[132] Agencies like this play an important role in ensuring there are legal consequences for users who post illegal content. While specialist agencies exist to monitor specific types of illegal content, having one dedicated agency for all illegal content on social media is a difficult task. Pulling out the areas of government connected with social media and internet platforms would not be easy. Social media has permeated most industries and new uses of the internet continue to disrupt the way people work and live.

D Rights or principles-based regulation

Discussions around social media and rights are multi-faceted. As large social media and other online platforms like Facebook and Google offer huge access to public life as gatekeepers, there is an idea that they are public spaces and users should have a right to be there and to be safe.[133] Access rights becomes complicated when framed as users having rights to use the platforms safely without exposure to harmful content versus rights of users

[128] Above.

[129] Every-Palmer, above n 22, at 18.

[130] *Online Harms White Paper*, above n 116, at 53

[131] Land, above n 105, at 380.

[132] National Center for Missing & Exploited Children <www.missingkids.org>

[133] Every-Palmer, above n 22, at 11.

not to be banned or have content removed without due process. There is a case for treating global social media platforms as more than private companies if companies are acting as state censors either by delegated responsibility or of their own initiative.[134] An international human rights model would provide a global standard but would require clarity on whether individual states enforce activity in their own jurisdictions.[135] Application would have to depend on the size of the user base and global power of the platform to put it above a small single market internet company.

Similar to framing regulation around rights, principles for a safe online environment could guide regulation. Murray is disappointed that the UK is going in a duty of care direction rather than in a principles direction. Principles have the potential for greater flexibility because they can provide high level expectations in legislation while details can be specified in guidelines and standards which are easier to change.[136] The Helen Clark Foundation presented the "Christchurch Principles" at the Paris Peace Conference in November 2019. The Christchurch Principles were designed to complement the Christchurch Call with the aim to have shared rights and responsibilities between technology companies, governments and society.[137] The Christchurch Principles are aimed at addressing wider online harms than terrorist and violent extremist content with objectionable content as a "tip of the iceberg" built up by social discord, failure to address "fake news" and radicalisation.[138] The principles are as follows:[139]

- Equal participation. Governments should ensure participation is not prevented by rights-violating content.
- Duty to protect. Governments should protect human rights and democratic norms.
- Responsibility to respect. Companies should respect rights either directly or through their capacity to influence.
- Responsibility to remedy. Governments and companies should be remedy violations of rights or inabilities to exercise rights to participate in democratic society.
- Structural change. An all of system approach should address the negative impacts of technology.

---

[134] Land , above n 105, at 404.

[135] Above at 394.

[136] Murray, above n 27, at 6.

[137] The Helen Clark Foundation *The Christchurch Principles* (November 2019) <helenclark.foundation/our-impact/> at 6.

[138] Above at 7.

[139] Above at 8-9.

- Duty of care. Governments should have a duty to ensure any regulatory intervention respects rights and companies should have a duty to address the negative consequences of their products.
- Decentralisation. Power in the digital realm should be consistent with democratic outcomes.
- Inclusivity. Diverse voices and perspectives should be included in decisions about technology.
- Communicative action. Communication should be trustworthy.

The principles support a mixed model regulatory approach which acknowledges the positive features social media brings to society.[140] Errington emphasises equal participation at the heart of the principles and champions the need for softer regulatory tools alongside law enforcement powers.[141] David Hall, one of the designers of the principles, believes that platforms prioritise freedom of expression for profit motives over other rights which are essential to democracy.[142] The principles, while well-meaning, have not largely been taken up by governments and companies although there is some movement towards duties of care which are often equated with principles. New Zealand alone could not interfere with the design and structure of platforms.

E Risks to freedom of expression

As social media is a place for debates of political and public significance, equality of participation is essential. Freedom of expression cannot be just for some if it effectively silences others. The concerns in overseas based literature are focussed on a concept referred to as censorship creep. If social media platforms have a statutory duty to moderate content and the only accountability is reporting, there is risk of over moderation and removal of content without giving a right of reply. Macdonald et al. are concerned that moral based regulation could silence activists in oppressive regimes from speaking against their governments and remove social media as a valuable tool for activism.[143] YouTube automatically removed content depicting atrocities in the war in Syria which could have been used in legitimate journalism to raise awareness of human rights abuses.[144] In the first reading of the Bill, multiple members mentioned the video of the killing of George Floyd

[140] Above at 10.

[141] Errington, above n 107, at [43].

[142] David Hall "How the Christchurch Principles will fight the spread of hate" (12 November 2019, The Spinoff) <www.thespinoff.co.nz>.

[143] Macdonald et al, above n 108, at 189.

[144] Above at 190.

by police in the US as an example of a video with a genuine public interest in being seen that was at risk of removal.[145] Under a duty of care or delegated censorship, the default response to questionable content could be removal.

Using take-down powers, New Zealand authorities are not likely to seek to compel social media platforms to silence genuine activists and those bringing attention to injustices. It is not intended that the online postings of news outlets on social media platforms will be subject to take-down notices in New Zealand for reporting on content that is in the public interest.[146] Nevertheless, it is important to raise these questions and give the law proper scrutiny for unintended consequences. There are real risks in creating a culture of over-blocking when delegating censorship powers and creating duties of care. Further evidence is needed before extending the law in that direction.

VII Comparative examples: Germany and Australia

A Germany and NetzDG

Germany's Netzwerkdurchsetzungsgesetz or Network Enforcement Act, known as Netz DG was passed in 2017 and has been called the most advanced country in the world for regulation of social media. Germany has taken a strong regulatory position. Jacinda Ardern has acknowledged the influence of Netz DG particularly on Facebook which has increased staffing as a result.[147] Netz DG applies to social media platforms with over 2 million users in Germany. Platforms are given a 24-hour period to take down manifestly unlawful content after receiving a complaint from either users or authorities. The 24-hour period is considered reasonable to press the need for fast action but also gives the online platform sufficient opportunity.[148] If platforms fail to comply, they can be fined up to 50 million euro. Platforms are required to report every six months about complaints received. NetzDG was amended in June 2021 to increase the information required in transparency reports, to make complaints processes easier to use, to require appeals mechanisms and to include video sharing platforms.[149] The lack of appeals process was previously criticised for the

[145] (11 February 2021) 749 NZPD (Films, Videos, and Publications Classification (Urgent Interim Classification of Publications and Prevention of Online Harm) Amendment Bill — First Reading).

[146] Regulatory Impact Statement, above n 36.

[147] Sam Sachdeva "How the world is tackling social media regulation" (Newsroom, 11 April 2019) <www.newsroom.co.nz>.

[148] Rochefort, above n 13, at 245.

[149] Library of Congress "Germany: Network Enforcement Act Amended to Better Fight Online Hate Speech" (2021) <www.loc.gov>.

lack of recourse to lessen over-blocking and limiting freedom of expression.[150] NetzDG is a form of limited or co-regulation as it relies on platforms to take measures to implement the law. NetzDG is does not place liability on platforms for their decisions on content, rather to comply platforms need to implement the procedural requirements.[151] The burden of compliance is placed upon platforms and the government monitors the administrative performance.[152]

The requirement to act on user complaints is much wider than New Zealand's proposed take-notices issued by Inspectors. Facebook has a two-step process for reviewing NetzDG reports, firstly against Facebook's Community Standards and then against the German Criminal Code.[153] Facebook assures that reported content is reviewed by trained professionals made up of employees, contractors and partner company staff and that where the legality is unclear it is referred to in-house lawyers or external counsel.[154] Facebook's transparency report for the period January to June 2021 reported 77,671 reports for 67,028 unique pieces of content mostly reported by individuals.[155] The report shows that most content is deleted or blocked at the Community Standards review stage, of 11,699 pieces of deleted or blocked content, only 1,092 were blocked because they violated the German Criminal Code but not Community Standards. Content is over-reported by users, but the number of blocked and removed content does not lead to a conclusion of automatic or excessive blocking. Reporting has increased significantly in 2021, with reports from previous periods from 2018-2021 ranging between 500 and 4,000.[156] The reporting volumes on Instagram have seen a similar increase in 2021.[157] Increased accessibility of the reporting form may explain the rise as previously the form was difficult to find so users were reporting through the normal community standards reporting tool.[158] Facebook was fined $2.3 million euro in 2019 for underreporting of complaints.[159] Germany's Federal

[150] Rochefort, above n 13, at 246.

[151] Thomas Kasakowskij, Julia Fürst, Jan Fischer, Kaja J. Fietkiewicz "Network enforcement as denunciation endorsement? A critical study on legal enforcement in social media" Telematics and Informatics 46 (2020) at 3.

[152] Rochefort, above n 13, at 246-247

[153] Facebook *NetzDG Transparency Report July 2021* (2021) <www.about.fb.com> at 4.

[154] Above at 7.

[155] Above at 5.

[156] Facebook "Where can I see Facebook's NetzDG Transparency Reports?" <www.facebook.com/help>.

[157] Instagram "Network Enforcement Act ("NetzDG") <help.instagram.com>.

[158] Thomas Escritt "Germany fines Facebook for under-reporting complaints" (Reuters, 3 July 2019) <www.reuters.com>.

[159] Deutsche Welle "Germany Fines Facebook for Underreporting Hate Speech Complaints" (2 July 2019) <www.dw.com>.

Office of Justice issued the penalty for incomplete reporting of complaints by not counting all categories of complaint.

NetzDG is not without controversy and criticism. It relies on informers or user-led surveillance rather than direct intervention by a regulator. Much of the criticism has been around the potential for over-blocking of content, however, Kasakowskii et al are concerned about the lack of accountability for over-blocking. Much of the criticism is around the enforcement of hate speech laws. Reporting tools can be used by political opponents to wrongly accuse other groups which can put freedom of speech and democracy at risk from both the left and the right. [160] Kasaowskii et al refer to research suggesting that over-blocking is not common and that platform operators are not acting differently to before NetzDG.[161]

B Australia and Abhorrent Violent Material

In response to 15 March 2019, Australia quickly passed the Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019 (AVM) in April 2019. The legislation was rushed through both houses of parliament in two days without any consultation.[162] The new offences are for failure to report AVM to the Australian Federal Police in a reasonable time regarding abhorrent violent conduct in Australia[163] and failure to ensure the expeditious removal of AVM from anywhere in the world where material is reasonably capable of being accessed in Australia.[164] Platforms can be fined up to $10,500,000 or 10% of their annual profit. The Act contains defence which cover off some of the major freedom of speech concerns including news reporting in the public interest, artistic work done in good faith and academic research.[165] Australia has an eSafety Commissioner who receives complaints but does not actively monitor the internet. The eSafety Commissioner can issue a notice to any website or web hosting service where AVM is published. The AVM notices are not "take-down" notices, however if platforms are prosecuted for hosting AVM, the notice can be used as evidence of recklessness The Australian Federal Police are still the main prosecuting agency.[166] As at 24 March 2020, the eSafety Commissioner had issued

[160] Kasakowskij et al, above n 149, at 4.

[161] Above

[162] Evelyn Douek "Australia's "Abhorrent Violent Material" Law: Shouting "Nerd. Harder" and Drowning Out Speech" (2020) 94 Austl. L.J. 41, at 2.

[163] Section 474.33

[164] Section 474.34

[165] Section 474.37

[166] eSafetyCommissioner "Abhorrent Violent Material: facts and falsehoods" (24 March 2020) <www.esafety.gov.au> at 1.

18 notices for what is described as "worst-of-the worst underground gore sites" showing beheadings, shootings and other murders.[167] The eSafety Commissioner is adamant that Australia's approach is not heavy handed and reflects community standards. The AVM contains limits and defences with a burden of proof on the defendant for example if AVM was published for legitimate journalism, academic or artistic purposes and only captures footage by perpetrators and associates not innocent bystanders or civilian journalists.[168]

The AVM was initially met with industry criticism due to the lack of industry consultation and the onerous requirements for platforms.[169] As with other duty of care models, there is concern about the realistic ability of platform operators to comply due to the enormous volume of user-generated content uploaded to social media. Similar concerns have been raised about NetzDG, that platforms may over block content out of fear of liability.[170] While the AVM is being used in a limited scope for the worst of the worst content, it has an extremely broad range applying to potentially any social media platform in the world.[171] While the Act contains seemingly adequate defences, in practice it is going to be near impossible for an algorithm or human moderator to be trained to accurately determine the context.[172] It also does not have equivalent transparency requirements to NetzDG. As Douek puts it, the legislation is essentially asking technology companies to "nerd harder" to create to the perfect enforcement tools without an appreciation of what is currently possible.[173] The legislation was rushed, lacked consultation and can rightly be criticised a not being well thought through, however so far it is not being enforced in an irresponsible way. The intention is to have tools to respond to the abhorrent events like 15 March 2019 that result in viral online content. New Zealand chose not to make urgent legislation[174] as it did for firearms and has taken the normal path towards passing a bill allowing for submissions which have been taken into account by the Select Committee.

---

[167] Above at 2.

[168] Valcic, above n 91, at 16.

[169] Above at 17.

[170] Above at 32.

[171] Douek, above n 160, at 4.

[172] Above at 10.

[173] Above at 12.

[174] Jenna Lynch, 'Jacinda Ardern Will Not Follow Australia's Hard-Line Response to Extremist Content', *Newshub* (19 July 2019) <www.newshub.co.nz>.

VIII Has New Zealand got it right?

New Zealand has the right focus in targeting the hosting of objectionable content which is the most urgent gap in the law. Regulating social media platforms generally in a broad sense is not necessary if the focus is to be on the most harmful content. Only addressing the hosting of illegal content through take-down notices may be open to criticism as being the tip of the iceberg but to delve into other social issues caused by the practices of social media companies is not without risks. The Workshop research concluded that regulatory approaches to social media are all essentially experimental so without a significant volume of evidence, regulators need to have the tools to be agile and responsive.[175] New Zealand will be contemplating the need for further regulation of harmful online content in a review of media and online content regulation announced in June 2021.[176]

New Zealand has not placed a duty of care on platforms to develop tools to limit objectionable content or introduced requirements for algorithm transparency. As the Australian and German examples have shown, it is possible to put the compliance burden back on the industry. This should be considered at a later time when there is a larger body of evidence about the effectiveness of overseas duty of care regimes. Relying on take-down powers alone could be called short sighted and reactive but amending the FVCP Act is not the end of the conversation. Requiring companies to take steps to keep users safe takes some of the burden off users to keep themselves safe online. As noted in the UK Government response to consultation on the White Paper, regulation is only one part of the solution. Governments need to support the development of technology to incorporate safety by design and empower and educate users to think critically.[177] The problem with duties of care is that the technology is not yet ready for it. It is currently too risk and opens up potential for misunderstanding and over-blocking.

A challenge is enforcement against the small to medium platforms that do not have the same resources and reputational responsibilities as the giants. Making contacts, taking an educative approach and requesting voluntary removal of content is an approach largely taken by New Zealand's Department of Internal Affairs and Australia's eSafety Commissioner.

---

[175] The Workshop, above n 14, at 23.

[176] Hon Jan Tinetti "Govt acts to protect NZers from harmful content" (10 June 2021) <www.beehive.govt.nz>.

[177] Department for Digital, Culture, Media & Sport and Home Office *Online Harms White Paper: Full government response to the consultation* (2020) at [18].

The maximum penalty for non-compliance with a take-down order is not enough to deter a large multination internet platform. The Helen Clark Foundation supports the idea of penalties based upon annual revenue like the Australian example.[178] As internet companies are largely commercially driven, the incentive to comply with the law needs to be linked to profits. Linking the maximum penalty to a percentage of the company's annual turnover as Australia has done would provide for proportionate penalties tailored to the platform size. Determining the appropriate level of regulation requires a balancing exercise. New Zealand is a small economy with a low user base on a global scale. Regulatory overreach may risk New Zealanders losing the benefits that these gateway platforms provide.

IX Conclusions

Social media platforms will always come up with new tools which may fall outside of regulation, the law always be behind. The addition of take-down notices and the removal of safe-harbour are a good start and will strike a balance between going too far and not going far enough in holding platforms to account. Relying on take-down notices is not a perfect system and could result in regulatory whack-a-mole next time an objectionable publication goes viral online. The difference that this sort of legislation will make remains to be seen. Enforcement will be rare and difficult. The value is in sending a message to the industry that governments can step in and hold platforms to account when they need to. Further reaching regulation may become necessary but should not be introduced in a hurry. New Zealand can go further in creating legal responsibilities for internet platforms to keep New Zealand users safe but must proceed with caution. New Zealand needs to keep an open mind to further regulating social media in future, once there is an available and reliable body of evidence about how effective overseas duty of care regimes are. The wider harms caused by online platforms should be closely monitored and future opportunities for regulation explored. At this point in time, there is much speculation on how such laws could be misused to infringe on rights and freedoms. Only addressing objectionable content at this stage deals with the most urgent need and keeping the censorship role with government retains control.

[178] Mason and Errington, above n 3, at 17.

BIBLIOGRAPHY

I PRIMARY SOURCES

A Legislation

1 New Zealand

Anti-Money Laundering and Countering Financing of Terrorism Act 2009

Companies Act 1993

Contempt of Court Act 2019

Fair Trading Act 1986

Films, Videos and Publications Classification Act 1993

Films, Videos and Publications Classification (Urgent Interim Classification of Publications and Prevention of Online Harm) Amendment Bill 2020 (268-2)

Harmful Digital Communications Act 2015

Health and Safety at Work Act 2015

High Court Rules 2016

Privacy Act 2020

2 Australia

Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019

3 Germany

Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken / Act to Improve Enforcement of the Law in Social Networks (Network Enforcement Act) 2017

4 United Kingdom

Draft Online Safety Bill 2021 (CP 405)

5 United States

Communications Decency Act 1996

B International instruments

Council of Europe, Convention on Cybercrime, 23 November 2001 (Budapest Convention)

C Case law

R v Arps [2019] NZDC 11547

Commerce Commission v Viagogo AG [2019] NZCA 472

Poynter v Commerce Commission SC 32/2009 [2010] NZSC 38

II SECONDARY SOURCES

A Books and chapters in books

Andrew Chen (ed.) Shouting Zeros and Ones: Digital Technology, Ethics and Policy in New Zealand (2020, Bridget Williams Books, online ed.)

Chapter 1: Curtis R. Barnes, Tom Barraclough "Digitised lies: New Zealand and the globalised disinformation threat";

Chapter 2: Katherine Errington, in conversation with Anjum Rahman "Reducing online harm"

B Journal articles

Raphael Cohen-Almagor "The Role of Internet Intermediaries in Tackling Terrorism Online" 86:2 Fordham Law Review 425-454

Evelyn Douek "Australia's "Abhorrent Violent Material" Law: Shouting "Nerd. Harder" and Drowning Out Speech" (2020) 94 Austl. L.J. 41, 50 n.77

John Ip "Law's response to New Zealand's 'darkest of days'" (2021) 50 Common Law World Review 1 21–37

Thomas Kasakowskij, Julia Fürst, Jan Fischer, Kaja J. Fietkiewicz "Network enforcement as denunciation endorsement? A critical study on legal enforcement in social media" (2020) Telematics and Informatics 46 101317

Molly K. Land "Against Privatized Censoship: Proposals for Responsible Delegation" (2020) 60:2 Virginia Journal of International Law 363

Stuart Macdonald, Sara Giro Correia and Amy-Louise Watkin "Regulating terrorist content on social media: automation and the rule of law" (2019) International Journal of Law in Context 15, 183–197

Andrew Murray "Rethinking Regulation for the Digital Environment" (2019) London School of Economics and Political Science, Policy Briefing 41

Alex Rochefort "Regulating Social Media Platforms: A Comparative Policy Analysis" (2020) Communication Law and Policy, 25:2, 225-260

Michelle Roter "With Great Power Comes Great Responsibility: Imposing a Duty to Take down Terrorist Incitement on Social Media" (2017) 45 Hofstra Law Review 1379

Peter A. Thompson "Beware of Geeks Bearing Gifts: Assessing the Regulatory Response to the Christchurch Call" (2019) 7 The Political Economy of Communication 1, 83–104

Alexander Tsesis, "Social Media Accountability for Terrorist Propaganda" (2017) 86 FORDHAM L. REV. 605

Jasmine Valcic "The Sharing of Abhorrent Violent Material Act: The Realities and Implications of Australia's New Laws Regulating Social Media Companies" (2021) 33 Bond Law Review 1 11-35

Lorna Woods "The duty of care in the Online Harms White Paper" (2019) 11 Journal of Media Law 1, 6-17

C Official reports

Department for Digital, Culture, Media and Sport, and the Home Office *Online Harms White Paper* (2019)

Department for Digital, Culture, Media & Sport and Home Office *Online Harms White Paper: Full government response to the consultation* (2020)

Department of Internal Affairs *Departmental disclosure - Films, Videos and Publications Classification (Urgent Interim Classification of Publications and Prevention of Online Harm) Amendment Bill*

Department of the Prime Minister and Cabinet and Ministry of Justice *Cabinet Paper: Budapest Convention on Cybercrime: Approval to Initiate the First Stage Towards Accession* (2020)

Royal Commission of Inquiry into the Terrorist Attack on Christchurch Mosques on 15 March 2019 *Hate speech and hate crime related legislation* (26 November 2020)

Royal Commission of Inquiry into the Attack on Christchurch Mosques on 15 March 2019 *Ko tō tātou kāinga tēnei Report: Royal Commission of Inquiry into the terrorist attack on Christchurch masjidain on 15 March 2019* (8 December 2020)

Treasury *Regulatory Impact Statement - Films, Videos and Publications Classification (Urgent Interim Classification of Publications and Prevention of Online Harm) Amendment Bill* (2020)

D Other reports

James Every-Palmer *Regulation of new technology: Institutions and processes* (The Law Foundation New Zealand, 2018)

Orla Lynskey *Regulating 'Platform Power'* (LSE Law, Society and Economy Working Papers 1/2017)

Claire Mason and Katherine Errington *Anti-social media: Reducing the spread of harmful content on social media networks* (The Helen Clark Foundation, 2019)

The Helen Clark Foundation *The Christchurch Principles* (2019)

The Workshop *Digital Threats to Democracy* (2019)

E Internet materials

Rt Hon Jacinda Ardern "PM House Statement on Christchurch mosques terror attack" (19 March 2019) <www.beehive.govt.nz>

Christchurch Call <www.christchurchcall.com>

eSafetyCommissioner "Abhorrent Violent Material: facts and falsehoods" (24 March 2020) <www.esafety.gov.au>

Deutsche Welle "Germany Fines Facebook for Underreporting Hate Speech Complaints" (2 July 2019) <www.dw.com>

Henry Cooke "The disparate global locations that make up New Zealand Facebook - and your news feed" (Stuff, 27 March 2017) <www.stuff.co.nz>

Marc Daalder "Government to boost efforts against online extremism" (Newsroom, 14 October 2019) <www.newsroom.co.nz>

Marc Daalder "New bill comes with online takedown powers" 27 May 2020 <www.newsroom.co.nz>

John Edwards "Dwarfed by the digital giants, here's how we can make our voice heard" (The Spinoff, 31 October 2019) <www.thespinoff.co.nz>

Thomas Escritt "Germany fines Facebook for under-reporting complaints" (Reuters, 3 July 2019) <www.Reuters.com>

Facebook *NetzDG Transparency Report July 2021* (2021) <www.about.fb.com>

Facebook *Update on New Zealand* (18 March 2019) <about.fb.com>

Facebook "Where can I see Facebook's NetzDG Transparency Reports?" <www.facebook.com/help>

David Hall "How the Christchurch Principles will fight the spread of hate" (The Spinoff, 12 November 2019) <www.thespinoff.co.nz>

InternetNZ "To block or not to block Technical and policy considerations of Internet filtering" (September 2019) <www.internetnz.nz>

Instagram "Network Enforcement Act ("NetzDG") <help.instagram.com>

Legislation Design and Advisory Committee *Legislation Guidelines* (2018) <www.ldac.govt.nz>

Library of Congress "Germany: Network Enforcement Act Amended to Better Fight Online Hate Speech" (2021) <www.loc.gov>

Jenna Lynch, 'Jacinda Ardern Will Not Follow Australia's Hard-Line Response to Extremist Content', *Newshub* (19 July 2019) <www.newshub.co.nz>

Hon Tracey Martin *New Bill to counter violent extremism online* (Press Release, 26 May 2020 <www.beehive.govt.nz>

National Center for Missing & Exploited Children <www.missingkids.org>

Statement by Press Secretary Jen Psaki on the Occasion of the United States Joining the Christchurch Call to Action to Eliminate Terrorist and Violent Extremist Content Online <www.whitehouse.gov>

Radio New Zealand "Authorities move to take down 'atrocious' mosque attack material" (25 June 2021) <www.rnz.co.nz>

Anjum Rahman "A year on, the Christchurch Call must go beyond 'don't livestream mass murder'" (The Spinoff, 15 May 2020) <www.thespinoff.co.nz>

Sam Sachdeva "How the world is tackling social media regulation" (Newsroom, 11 April 2019) <www.newsroom.co.nz>

Sam Shead "Facebook owns the four most downloaded apps of the decade" (BBC News, 18 December 2019) <www.bbc.com>

Tech against terrorism *Analysis: New Zealand attack and the terrorist use of the internet* (26 March 2019) <www.techagainstterrorism.org>

Hon Jan Tinetti "Govt acts to protect NZers from harmful content" (10 June 2021) <www.beehive.govt.nz>

Laura Walters "Govt research: NZ importing right-wing extremist content" (Newsroom, 13 January 2021) <www.newsroom.co.nz>

Danny Yadron "Twitter deletes 125,000 Isis accounts and expands anti-terror teams" (The Guardian, 5 February 2016) <www.theguardian.com>