After Christchurch:
Hate, harm and the limits of censorship

4. Regulating harmful communication: Current legal frameworks

David Bromell

Working Paper 21/05

VICTORIA UNIVERSITY OF WELLINGTON TE HERENGA WAKA
Institute for Governance and Policy Studies
A research institute of the School of Government

CAS CENTER FOR ADVANCED INTERNET STUDIES

INSTITUTE FOR GOVERNANCE AND POLICY STUDIES
WORKING PAPER
21/05

MONTH/YEAR

March 2021

AUTHOR

David Bromell
Senior Associate
Institute for Governance and Policy Studies

INSTITUTE FOR GOVERNANCE AND POLICY STUDIES

School of Government
Victoria University of Wellington
PO Box 600
Wellington 6140
New Zealand

For any queries relating to this working paper, please contact igps@vuw.ac.nz

ACKNOWLEDGEMENT

Research on this series of working papers has been financially supported by a fellowship at the Center for Advanced Internet Studies (CAIS) in Bochum, NRW, Germany (Oct 2020—Mar 2021).

This is paper four in a series of seven working papers, **After Christchurch: Hate, harm and the limits of censorship.**

The series aims to stimulate debate among policy advisors, legislators and the public as New Zealand considers regulatory responses to 'hate speech' and terrorist and violent extremist content online following the terrorist attack on Christchurch mosques in March 2019 and the Royal Commission of Inquiry that reported in November 2020.

The seven working papers in this series are:

Title	Reference
1. The terrorist attack on Christchurch mosques and the Christchurch Call	WP 21/02
2. 'Hate speech': Defining the problem and some key terms	WP 21/03
3. Challenges in regulating online content	WP 21/04
4. Regulating harmful communication: Current legal frameworks	WP 21/05
5. Arguments for and against restricting freedom of expression	WP 21/06
6. Striking a fair balance when regulating harmful communication	WP 21/07
7. Counter-speech and civility as everyone's responsibility	WP 21/08

Dr David Bromell is currently (until March 31, 2021) a research Fellow at the Center for Advanced Internet Studies (CAIS) in Bochum, North Rhine-Westphalia, Germany, which has supported his research on this series of working papers. He is a Senior Associate of the Institute for Governance and Policy Studies in the School of Government at Victoria University of Wellington, and a Senior Adjunct Fellow in the Department of Political Science and International Relations at the University of Canterbury. From 2003 to 2020 he worked in senior policy analysis and advice roles in central and local government.

He has published two monographs in Springer's professional book series:

- *The art and craft of policy advising: A practical guide* (2017)
- *Ethical competencies for public leadership: Pluralist democratic politics in practice* (2019).

Contents

Regulating harmful communication: Current legal frameworks

Abstract

Previous working papers in this series have discussed the terrorist attack on Christchurch mosques in March 2019 and the subsequent Christchurch Call to eliminate terrorist and violent extremist content online (Working paper 21/02), defined 'hate speech' and some other key terms (Working paper 21/03), and outlined challenges in regulating online content (Working paper 21/04).

This paper summarises provisions in international human rights standards, and in New Zealand law, that protect and qualify the right to freedom of expression. It also notes relevant recommendations of the Royal Commission of Inquiry into the terrorist attack on Christchurch mosques. The paper then summarises regulation of harmful communication and some recent developments in the UK, Australia, Canada, Germany, France and the European Union.

A free, open and democratic society protects everyone's right to freedom of opinion and expression but may justifiably qualify this right to prevent harm to others, if it does so in ways that conform to strict tests of legality, proportionality and necessity. There is an established consensus in international human rights law that it may be justifiable to restrict public communication that incites discrimination, hostility or violence against a social group with a common 'protected characteristic' such as nationality, race or religion.

Regulation to protect people from criticism, offence or lack of respect is not a justifiable restriction of freedom of expression, and international human rights standards specifically discourage blasphemy laws. The recommendation of the Royal Commission of Inquiry that religion become a 'protected characteristic' *without qualification* in New Zealand's regulation of harmful communication risks departing from an established consensus in international human rights law.

International human rights standards also urge careful distinctions between (1) communication that is illegal and a criminal offence; (2) communication that is not a criminal offence but may justify a civil suit; and (3) so-called 'lawful hate speech' that does not give rise to criminal or civil sanctions, but may still raise concerns about tolerance, civility and respect for others.

The remaining three working papers in this series explore arguments for and against restricting freedom of expression (Working paper 21/06), the need to strike a fair balance when regulating harmful communication (Working paper 21/07), counter-speech strategies as alternatives or complements to prohibition and censorship, and civility as everyone's responsibility (Working paper 21/08).

Tags: #ChristchurchCall #ChristchurchAttack #hatespeech #freespeech #censorship

Introduction: A qualified right to freedom of expression

A free, open and democratic society protects everyone's right to freedom of opinion and expression but may justifiably qualify this freedom to prevent harm to others, if it does so in ways that conform to strict tests of legality, proportionality and necessity.

Principles for regulating justifiable restrictions on freedom of expression are well established in international human rights standards. It is useful to review how these principles currently apply in New Zealand law and in a number of other jurisdictions before considering whether, when and how to amend existing regulation, as recommended by the Royal Commission of Inquiry into the terrorist attack on Christchurch mosques (Royal Commission of Inquiry, 2020a, 2020b).

Freedom of expression in international human rights law

The right to freedom of expression

Article 19 of the Universal Declaration of Human Rights affirms the right to freedom of opinion and expression:

> Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers (UN General Assembly, 1948).

Article 19(1)(2) of the International Covenant on Civil and Political Rights (ICCPR) elaborates on this and requires states to guarantee this freedom to all people, while noting that it is a *qualified freedom*—qualified, that is, by Article 19(3) and Article 20:

> **Article 19**
>
> 1. Everyone shall have the right to hold opinions without interference.
> 2. Everyone shall have the right to freedom of expression; this right shall include freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his [sic.] choice.
> 3. The exercise of the rights provided for in paragraph 2 of this article carries with it special duties and responsibilities. It may therefore be subject to certain restrictions, but these shall only be such as are provided by law and are necessary:
>   (a) For respect of the rights or reputations of others;
>   (b) For the protection of national security or of public order (*ordre public*), or of public health or morals.
>
> **Article 20**
>
> 1. Any propaganda for war shall be prohibited by law.
> 2. Any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law (UN General Assembly, 1966).

A Human Rights Committee General Comment on Article 19 states: 'The scope of paragraph 2 embraces even expression that may be regarded as deeply offensive, although such expression may be restricted in accordance with the provisions of article 19, paragraph 3 and article 20' (UN Human Rights Committee, 2011, para. 11).

Restrictions on freedom of expression

Because freedom of expression is a qualified right, it may be justifiable to restrict it in certain circumstances.

Relevant provisions are the ICCPR, the International Convention on the Elimination of All Forms of Racial Discrimination 1965 (ICERD), and the Convention on the Prevention and Punishment of the Crime of Genocide 1948 (Genocide Convention). New Zealand has ratified all three conventions, which set out restrictions that states *must* or *may* place on freedom of expression.

International Covenant on Civil and Political Rights

The three protected characteristics under article 20(2) of the ICCPR (nationality, race and religion) have been interpreted as supporting a broader principle of equality and protection from harmful communication (Human Rights Commission, 2019, p. 13). UN Special Rapporteur David Kaye has, for example, argued that:

> United Nations human rights standards offer broader protection against discrimination than that afforded through the focus in article 20 (2) on national, racial or religious hatred … Given the expansion of protection worldwide, the prohibition of incitement should be understood to apply to the broader categories now covered under international human rights law (UN General Assembly, 2019, para. 9).

The Special Rapporteur immediately added, however:

> A critical point is that the individual whose expression is to be prohibited under article 20(2) of the Covenant is the advocate whose advocacy constitutes incitement. A person who is not advocating hatred that constitutes incitement to discrimination, hostility or violence, for example, a person advocating a minority or even offensive interpretation of a religious tenet or historical event, or a person sharing examples of hatred and incitement to report on or raise awareness of the issue, is not to be silenced under article 20 (or any other provision of human rights law). Such expression is to be protected by the State, even if the State disagrees with or is offended by the expression. There is no 'heckler's veto' in international human rights law (UN General Assembly, 2019, paras 10–11).[1]

[1] On the 'heckler's veto', Kaye cites Aswad (2013, p. 1322):

> The stated general animating purpose behind Article 20(2) [of the ICCPR] was that, in the aftermath of the Nazi atrocities, it was viewed as necessary to proscribe speech that was intended to and would lead to such atrocities. Article 20(2) was included to prohibit advocacy inciting harm against a targeted national, racial, or religious group. In other words, the point of Article 20(2) was to prohibit expression where the speaker intended for his or her speech to cause hate in listeners who would agree with the hateful message and therefore engage in harmful acts towards the targeted group. There is no indication in the negotiating history that Article 20 was intended to prohibit speech about a targeted group that would offend the feelings of members of that group. There is certainly no suggestion in the negotiating history that speech should be banned if the targeted group would take offense to or oppose the message and members of the group display rejection of the message through violence or other harmful acts against the speaker or those associated with the speaker. In short, Article 20(2) was not meant to embody in human rights law a "heckler's veto," which would mandate the stifling of speakers when those who are offended choose to show their displeasure through harmful acts.

The UN Human Rights Committee has clarified that ICCPR article 20 does not necessarily require that 'hate speech' be made a criminal, as opposed to a civil offence, and that any prohibition under article 20 must also comply with article 19(3) and 'conform to the strict tests of necessity and proportionality' (UN Human Rights Committee, 2017, para. 10.4).

Seventeen states, including New Zealand, have entered 'reservations' to article 20 of the ICCPR. New Zealand's reservation reads:

> The Government of New Zealand having legislated in the areas of the advocacy of national and racial hatred and the exciting of hostility or ill will against any group of persons, and having regard to the right of freedom of speech, reserves the right not to introduce further legislation with regard to article 20 (UN OHCHR, n.d.).

International Convention on the Elimination of all Forms of Racial Discrimination

Article 4 of ICERD (UN General Assembly, 1965) requires states to make racially motivated 'hate speech' an offence:

> States Parties condemn all propaganda and all organizations which are based on ideas or theories of superiority of one race or group of persons of one colour or ethnic origin, or which attempt to justify or promote racial hatred and discrimination in any form, and undertake to adopt immediate and positive measures designed to eradicate all incitement to, or acts of, such discrimination and, to this end, with due regard to the principles embodied in the Universal Declaration of Human Rights and the rights expressly set forth in article 5 of this Convention, inter alia:
> (a) Shall declare an offence punishable by law all dissemination of ideas based on racial superiority or hatred, incitement to racial discrimination, as well as all acts of violence or incitement to such acts against any race or group of persons of another colour or ethnic origin, and also the provision of any assistance to racist activities, including the financing thereof;
> (b) Shall declare illegal and prohibit organizations, and also organized and all other propaganda activities, which promote and incite racial discrimination, and shall recognize participation in such organizations or activities as an offence punishable by law;
> (c) Shall not permit public authorities or public institutions, national or local, to promote or incite racial discrimination.

Unlike the ICCPR, article 4 does require criminalisation of racially motivated 'hate speech', including 'dissemination of ideas' and not only incitement to discrimination and acts of violence. The UN Committee on the Elimination of Racial Discrimination (2013, para. 13) recommended that states 'declare and effectively sanction as offences punishable by law:

> (a) All dissemination of ideas based on racial or ethnic superiority or hatred, by whatever means;
> (b) Incitement to hatred, contempt or discrimination against members of a group on grounds of their race, colour, descent, or national or ethnic origin;
> (c) Threats or incitement to violence against persons or groups on the grounds in (b) above;
> (d) Expression of insults, ridicule or slander of persons or groups or justification of hatred, contempt or discrimination on the grounds in (b) above, when it clearly amounts to incitement to hatred or discrimination;
> (e) Participation in organizations and activities which promote and incite racial discrimination.'

The Committee recommended, however, that the following contextual factors be taken into account:

- The content and form of speech, and the style in which it is delivered;
- The economic, social and political climate prevalent at the time the speech was made and disseminated;
- The position or status of the speaker in society and the audience to which the speech is directed;
- The reach of the speech, including the nature of the audience and the means of transmission; and
- The objectives of the speech (para. 15).

The Committee has further recommended that criminalisation of forms of racist expression should be reserved for serious cases, to be proven beyond reasonable doubt (para. 12). The Committee cautioned against broad or vague restrictions on freedom of speech and stressed that 'measures to monitor and combat racist speech should not be used as a pretext to curtail expressions of protest at injustice, social discontent or opposition' (para. 20).

'Lawful hate speech'

A 2011 report by Frank La Rue, Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, urged clear distinctions between three categories of expression:

> (a) Expression that constitutes an offence under international law and can be prosecuted criminally; (b) expression that is not criminally punishable but may justify a restriction and a civil suit; and (c) expression that does not give rise to criminal or civil sanctions, but still raises concerns in terms of tolerance, civility and respect for others. These different categories of content pose different issues of principle and call for different legal and technological responses (UN General Assembly, 2011b, para. 18).

The third category (c) is often referred to as 'lawful hate speech'.[2] The Special Rapporteur reiterated these distinctions in his report the following year (UN General Assembly, 2012a). They have been broadly affirmed in international human rights law (Human Rights Commission, 2019, p. 15).

Defamation of religion

Between 1999 and 2010, the UN voted on and adopted several non-binding resolutions on 'defamation of religion'. The motions were sponsored on behalf of the Organisation of the Islamic Conference, currently known as the Organisation of Islamic Co-operation (OIC). Opinion was split on these proposals between 2001 and 2010, which were opposed in many Western democracies and by human rights and free speech advocates and other religious groups.

In 2011, the OIC changed its approach and the UN Human Rights Council unanimously adopted Resolution 16/18: *Combating intolerance, negative stereotyping and stigmatization of, and discrimination, incitement to violence and violence against, persons based on religion or belief* (UN General Assembly, 2011a).

[2] Kwame Anthony Appiah (2021) commented in a recent column: 'Many of the most hurtful, cruel and despicable things people do are perfectly lawful. (And some unlawful things are perfectly harmless.)'

The UN Human Rights Committee followed this in July 2011 with the adoption of General Comment 34 on the ICCPR, which includes the statement that:

> Prohibitions of displays of lack of respect for a religion or other belief system, including blasphemy laws, are incompatible with the Covenant, except in the specific circumstances envisaged in article 20, paragraph 2, of the Covenant. Such prohibitions must also comply with the strict requirements of article 19, paragraph 3, as well as such articles as 2, 5, 17, 18 and 26. Thus, for instance, it would be impermissible for any such laws to discriminate in favour of or against one or certain religions or belief systems, or their adherents over another, or religious believers over non-believers. Nor would it be permissible for such prohibitions to be used to prevent or punish criticism of religious leaders or commentary on religious doctrine and tenets of faith (UN Human Rights Committee, 2011, para. 48).

The General Comment further states (at paragraph 49):

> Laws that penalize the expression of opinions about historical facts are incompatible with the obligations that the Covenant imposes on States parties in relation to the respect for freedom of opinion and expression. The Covenant does not permit general prohibition of expressions of an erroneous opinion or an incorrect interpretation of past events. Restrictions on the right of freedom of opinion should never be imposed and, with regard to freedom of expression, they should not go beyond what is permitted in paragraph 3 or required under article 20.

The UN General Assembly (2012b) welcomed and affirmed the Human Rights Council's Resolution 16/18. These resolutions were reaffirmed by the Human Rights Council in April 2013 (UN Human Rights Council, 2013).

The Rabat Plan

In 2011, the United Nations Office of the High Commissioner for Human Rights (OHCHR) held regional expert workshops in Vienna, Nairobi, Bangkok, Santiago and Rabat on incitement to national, racial or religious hatred under article 20 of the ICCPR. The workshops resulted in the adoption of the *Rabat Plan of Action on the prohibition of advocacy of national, racial or religious hostility or violence.*

In presenting the report, the OCHCR adopted La Rue's three categories of expression (UN General Assembly, 2011b, para. 18), urging 'a careful distinction between (a) forms of expression that should constitute a criminal offence; (b) forms of expression that are not criminally punishable, but may justify a civil suit; and (c) forms of expression that do not give rise to criminal or civil sanctions, but still raise concerns in terms of tolerance, civility and respect for the convictions of others' (UN General Assembly, 2013, Annex, para. 12).

The Rabat Plan sets a high threshold, even for harmful speech that incites hatred:

> Such threshold must take into account the provisions of article 19 of the Covenant. Indeed the three-part test (legality, proportionality and necessity) for restrictions also applies to cases involving incitement to hatred, in that such restrictions must be provided by law, be narrowly defined to serve a legitimate interest, and be necessary in a democratic society to protect that interest. This implies, among other things, that restrictions are clearly and narrowly defined and respond to a pressing social need; are the least intrusive measure available; are not overly broad, so that they do not restrict speech in a wide or untargeted way; and are proportionate so that the benefit to the protected interest outweighs the harm

> to freedom of expression, including with respect to the sanctions they authorize (UN General Assembly, 2013, Appendix, para. 18).

A six-part threshold test is proposed for expressions considered as criminal offences: content, speaker, intent, content and form, extent of the speech act, and likelihood (including imminence) of harm (Appendix, para. 29).

The Rabat Plan noted a need for robust definitions of key terms such as hatred, discrimination, violence and hostility, drawing on definitions and guidance provided in Article 19 of the Camden Principles on Freedom of Expression and Equality (Article 19, 2009).[3] The Rabat Plan further recommended (Appendix, para. 25) that 'states that have blasphemy laws should repeal them, as such laws have a stifling impact on the enjoyment of freedom of religion or belief, and healthy dialogue and debate about religion'. The OCHCR's report insists that:

> Restrictions must be formulated in a way that makes clear that its sole purpose is to protect individuals and communities belonging to ethnic, national or religious groups, holding specific beliefs or opinions, whether of a religious or other nature, from hostility, discrimination or violence, rather than to protect belief systems, religions or institutions as such from criticism. The right to freedom of expression implies that it should be possible to scrutinize, openly debate and criticize belief systems, opinions and institutions, including religious ones, as long as this does not advocate hatred that incites violence, hostility or discrimination against an individual or group of individuals (UN General Assembly, 2013, Annex, para. 11).

Recent developments

In 2019, Special Rapporteur David Kaye provided examples of expression that under international human rights standards should not be restricted:

- Lack of respect for a religion or other belief system, including blasphemy laws, except when advocacy of religious hatred constitutes incitement;
- Opinions that are 'erroneous' and 'an incorrect interpretation of past events' may not be subject to general prohibition—which calls into question laws that criminalise the denial of the Holocaust and other atrocities;
- A situation in which a speaker is 'individually targeting an identifiable victim' but not seeking to 'incite others to take an action against persons on the basis of a protected characteristic'; and
- Expression that may be offensive or characterised by prejudice and that may raise serious concerns of intolerance but may not meet a threshold of severity to merit any kind of restriction (UN General Assembly, 2019, paras 21–24).

He noted (para. 24): 'There is a range of expression of hatred, ugly as it is, that does not involve incitement or direct threat, such as declarations of prejudice against protected groups' but emphasised that:

> The absence of restriction does not mean the absence of action; States may (and should, consistent with Human Rights Council resolution 16/18) take robust steps, such as government condemnation of prejudice, education, training, public service announcements and community projects, to counter such intolerance and ensure that public authorities protect individuals against discrimination rooted in these kinds of assertions of hate (ibid.).

[3] See Working Paper 21/03, **'Hate speech': Defining the problem and some key terms**, p. 11, fn. 7.

In September 2019, a joint open letter signed by 26 United Nations experts raised concerns about the global increase in 'hate speech' and concluded: 'States should actively work towards policies that guarantee the rights to equality and non-discrimination and freedom of expression, as well as the right to live a life free of violence through the promotion of tolerance, diversity and pluralistic views; these are the centre of pluralistic and democratic societies' (UN OHCHR, 2019).

Context matters

In introducing their edited publication, *The content and context of hate speech*, Herz & Molnar (2012, p. 4) remind us that while international law sets some parameters and provides guidance, context matters:

> It is at least questionable whether it would be either possible or desirable to establish a standard global regulatory policy toward 'hate speech.' International law can be helpful in pushing for narrower restrictions on freedom of speech and thus reducing the risk of regulatory abuses, but one premise of almost every contribution to this collection is that there is no single means by which 'hate speech' can and should be addressed.

This is an important consideration in relation to denial of the Holocaust and other atrocities. Thomas Nagel (2002, p. 44) has argued:

> I think it is already sufficiently inexcusable that anyone should be jailed or fined for denying that the Holocaust took place or selling books that deny it or for conducting a mail order business in Nazi medallions, small busts of Hitler, and so forth. Those restrictions are deeply offensive in themselves, and I believe they are damaging to the situation of Jews in those societies that enforce them. They carry the message that the reality of the Holocaust and the evil of Nazism are propositions that cannot stand up on their own—that they are so vulnerable to denial that they need to be given the status of dogma, protected against criticism and held as articles of faith rather than reason.

Nadine Strossen (2018) and Kenan Malik (in Molnar, 2012, p. 87) argue similarly. Strossen (p. 23) quotes Anna Sauerbray, editor of German newspaper *Der Tagesspiegel*:

> The American way of dealing with Nazism ... always seemed to me the more mature way of handling threats to liberal democracy. Germany's [outlawing of 'hate speech'] seems like a permanent declaration of distrust in ... argument and ... education ... I have faith in a democratic public's ability to police itself. I wish Germany did.

On the other hand, David Fraser (2009, p. 520) argues that 'legal regulation of Holocaust denial … is best justified by a collective social, ethical, and political decision that Holocaust denial is an evil which we cannot permit or tolerate because the ethics of the Holocaust as perpetuated by deniers is the denial of all ethical possibilities within human existence.' The harm of Holocaust denial, he argues, is 'the *malum in se* of denial's attacks on our Western, twenty-first century memory and truth' (p. 537).

As Julie Suk (2012, p. 149) puts it:

> In France, as in Germany and other European nations, criminal law has played a critical role in enabling the state and the society to face their collective responsibility for the Holocaust. The criminal process can be described as a 'realm of memory,' through which the nation collectively remembers the historical circumstances that engendered its current constitutional commitments.

The European Union's 2008 Framework Decision 2008/913/JHA on combating certain forms and expressions of racism and xenophobia by means of criminal law sets out the obligation for member states to penalise:

- Incitement to violence or hatred directed against a group of persons or a member of such a group defined by race, colour, religion, descent or national or ethnic origin; and
- Public condoning, gross trivialisation or denial of the Holocaust (European Union, 2008).

Bhikhu Parekh (2012, p. 55) reminds us that:

> Context plays an important part in our assessment of the value of a law. Banning Holocaust denial has a particular meaning in Germany. It is part of reparative justice, a public statement of the country's acknowledgment of and apology for its past, a way of fighting neo-Nazi trends in German society, and so on. The ban is justified not only because Holocaust denial is a form of hate speech, but on other moral and historical grounds as well.

Sixteen European countries, as well as Israel, currently have laws against Holocaust denial and/or denial of genocide and/or denial of war crimes to incite discrimination.[4] A number of other countries, including Australia, have prosecuted Holocaust denial under broader laws that ban 'hate speech' or 'racial vilification'. It is debatable whether context justifies these restrictions in every case.[5]

Restrictions on freedom of expression in New Zealand law

The right to freedom of expression

The right to freedom of expression is protected in domestic law by the New Zealand Bill of Rights Act 1990 (NZBORA), section 14: 'Everyone has the right to freedom of expression, including the freedom to seek, receive, and impart information and opinions of any kind in any form.'

The Human Rights Commission (2019, p. 9) references *Handyside v UK* (5493/72) [1976] ECHR 5, para. 49, as cited by the New Zealand Court of Appeal in *Living Word Distributors Ltd v Human Rights Action Group* (2000, para. 45):

> Freedom of expression constitutes one of the essential foundations of a [democratic] society … it is applicable not only to 'information' or 'ideas' that are favourably received or regarded as inoffensive or as a matter of indifference, but also to those that offend, shock or disturb the State or any sector of the population. Such are the demands of that pluralism, tolerance and broadmindedness without which there is no 'democratic society'.

In New Zealand law, the right to freedom of expression is not, however, an unqualified freedom or what Feinberg (1973, pp. 86–87) calls an absolute or 'categorically exceptionless' right. NZBORA

[4] European countries with laws prohibiting Holocaust denial and/or genocide are: Germany, Austria, Belgium, the Czech Republic, France, Hungary, Italy, Liechtenstein, Lithuania, Luxembourg, Poland, Romania, Russia, Slovakia, Switzerland. Portugal prohibits denial of war crimes to incite discrimination. See further Schauer, 2012; and Suk, 2012.

[5] Apart from a visit by David Irving in the 1980s, New Zealand has relatively little history of Holocaust denial, but see Caldwell (2012) for a discussion of three cases at New Zealand universities that brought the issue to public attention: the Master of Arts theses of Joel Hayward and Roel van Leuween, and the potential PhD thesis of Hans Kupka.

section 5 states, however: 'the rights and freedoms contained in this Bill of Rights may be subject only to such reasonable limits prescribed by law as can be demonstrably justified in a free and democratic society.'

The test to be applied under section 5 was set by the Supreme Court in *Hansen v R*, which in turn drew on the decision of the Canadian Supreme Court in *R v Oakes.* The components of the so-called Hansen Test, set by Tipping J., are:

> (a) Does the limiting measure serve a purpose sufficiently important to justify curtailment of the right or freedom?
> (b) (i) Is the limiting measure rationally connected with its purpose?
> (ii) Does the limiting measure impair the right or freedom no more than is reasonably necessary for sufficient achievement of its purpose?
> (iii) Is the limit in due proportion to the importance of the objective? (*Hansen v R*, 2007, para. 104).

The Royal Commission of Inquiry (2020b, p. 5) noted that 'There is considerable scope for argument and controversy as to what are "reasonable limits" when it comes to the right to freedom of expression.'

Restrictions on freedom of expression

In New Zealand's commercial law, free speech is regulated in terms of copyright (Copyright Act 1994) and misrepresentation in contract (Contract and Commercial Law Act 2017). The law of defamation (Defamation Act 1992) creates a civil claim for damage to reputation from publishing untrue statements.

If the offending material has been published, complaints can be made to the publisher or broadcaster, as well as the Broadcasting Standards Authority, the Advertising Standards Authority or the New Zealand Press Council. It can also be submitted for classification to the Office for Film and Literature Classification under the Films, Videos, and Publications Classification Act 1993.[6] If the offence relates to a workplace, the Human Rights Commission (n.d.) advises that 'sometimes, the best action is for people to take their concern to the person's employer or to the CEO of the board or organisation where they work'.[7]

Under the Harmful Digital Communications Act 2015 s22, it is a criminal offence to post a digital communication that causes harm, intends to cause harm or would cause harm ('serious emotional distress') to 'an ordinary reasonable person', with perpetrators facing imprisonment for up to two years or a fine not exceeding $50,000. Protected characteristics are colour, race, ethnic or national origins, religion, gender, sexual orientation and disability (s6(1), Principle 10).

[6] New Zealand's Chief Censor used existing regulation to classify the Christchurch mosque shooter's manifesto and livestream video as 'objectionable', effectively banning their possession and distribution in New Zealand. Charges have been laid and successfully prosecuted, and New Zealand has twice succeeded in halting publication of the manifesto overseas, in Ukraine and Italy (Manch, 2020).

[7] In January 2021, the Employment Relations Authority dismissed Ivan Ilin's claim of unfair dismissal at milk processing facility Meadow Fresh owned by Goodman Fielder NZ Ltd when he was summarily dismissed after drawing a swastika on his overalls and allegedly saying 'so scary' to his colleagues on March 19, 2019, four days after the terrorist attack on Christchurch mosques (Dillane, 2021).

Threats of terrorist action are a crime (Terrorism Suppression Act 2002). Offensive behaviour or language in a public place is a summary offence, including addressing any words to any person intending to threaten, alarm, insult, or offend that person (Summary Offences Act 1981, s4).[8] It is also a summary offence to frighten or intimidate another person, including by threats of violence, or threats to unreasonably disrupt a public meeting (Summary Offences Act 1981, ss21, 37). Threats of harm to people and property are an offence under s307A of the Crimes Act 1961, with a penalty of up to seven years' imprisonment. A hate motivation for offences can be considered as an aggravating factor at sentencing (Sentencing Act 2002, s9(1)(h)).

Protection of freedom of expression in NZBORA is further qualified in New Zealand law by the Human Rights Act 1993. Under s131 (*Inciting racial disharmony*), it is a criminal offence to publish threatening, abusive or insulting material with the intention of exciting hostility or ill-will against, or bringing into contempt or ridicule, any group of persons in New Zealand on the ground of the colour, race, or ethnic or national origins of that group of persons. The Human Rights Act stipulates (s132) that the approval of the Attorney-General must be obtained before proceeding to a prosecution under s131.

Under s61 (*Racial disharmony*), it is a civil offence to publish or distribute, or use words in a public place, or use words knowing they are reasonably likely to be published or broadcast, that are threatening, abusive or insulting and likely to 'excite hostility against or bring into contempt any group of persons in or who may be coming to New Zealand on the ground of the colour, race, or ethnic or national origins of that group of persons.'

The criminal offence is punishable by conviction and a fine of up to NZD$7000, or up to three months' imprisonment. The civil prohibition is enforced by someone suing in the Human Rights Review Tribunal for civil damages (Edgeler, 2020).[9]

There have only been three decisions applying sections 61 and 131 of the Human Rights Act 1993 and equivalent provisions in earlier legislation. The three cases are discussed in the companion report of the Royal Commission of Inquiry (2020b, pp. 21–25). The most significant of these is the 2018 High Court judgment in *Wall v Fairfax NZ Ltd*, which demonstrated that 'liability under section 61 of the Human Rights Act is hard to establish, particularly once significant weight is afforded to the right to freedom of expression' (Royal Commission of Inquiry, 2020b, p. 24).

New Zealand law does not provide specifically for 'hate speech' in relation to religion, gender, disability, sexual orientation or other grounds, although such grounds can be considered as an aggravating factor under the Sentencing Act 2002 (Duff, 2019; Mason & Errington, 2019, p. 15). A

[8] In February 2021, Dean Rowe was convicted of offensive behaviour in the Dunedin District Court for racially abusing an Indian family at the Signal Hill lookout in December 2020 (Kidd, 2021).

[9] Based on a column he had written in the *National Business Review* in 2018, Renae Maihi started an online petition calling for Sir Bob Jones to be stripped of his knighthood. He sued her for defamation, accusing her of calling him a racist, writing 'hate speech' and calling for his knighthood to be revoked. When the case went to the High Court in Wellington in February 2020, Sir Bob discontinued defamation proceedings five days into a hearing set down for two weeks. This removed the court's responsibility to rule on the defamation charge and its opportunity to provide guidance on what counts as 'racism' and 'hate speech' (Black, 2020).

law against 'blasphemous libel' (Crimes Act 1961, s.123) was repealed by the Crimes Amendment Act 2019, which received royal assent just four days before the Christchurch mosque shootings.[10]

In August 2020, following the terrorist attack on Christchurch mosques, the Government introduced a Film, Video and Publications Classification (Urgent Interim Classification of Publications and Prevention of Online Harm) Amendment Bill with the objective of preventing and mitigating harms caused by objectionable publications. The Bill had its first reading on February 11, 2021 and was referred to the Governance and Administration Select Committee.

The Bill proposes to:

- Make livestreaming of objectionable content a criminal offence;
- Confer additional authority on the Chief Censor;
- Authorise an Inspector of Publications to issue a take-down notice for objectionable online content;
- Impose a civil pecuniary penalty on online content hosts that do not comply with an issued take-down notice in relation to objectionable online content;
- Provide that the 'safe harbour' provisions in the Harmful Digital Communications Act 2016 (HDC Act) which apply to online content hosts do not apply to the operation of the bill when enacted; and
- Facilitate the setting up of future government-backed mechanisms for blocking or filtering objectionable online content (Parliamentary Services, 2020)—that is, authorising the use of 'electronic systems' to prevent access to 'objectionable material' (a defined term in the principal Act).

According to the May 2020 Regulatory Impact Statement, the Christchurch mosque attacks exposed the following regulatory gaps in New Zealand's domestic law:

- Livestreaming objectionable content is not explicitly an offence against the [Film, Videos and Publications Classification] Act;
- The Chief Censor must publish a written decision within five working days of classifying a publication, which can delay initial decisions on objectionable content;
- The Department of Internal Affairs has no explicit power to request and/or enforce online content hosts to remove objectionable content from their platforms;
- The penalties under the Act and the deterrent factor they play are no longer appropriate in today's digital landscape, e.g., offences for non-compliance do not exist specifically for online content;
- The HDC 'safe harbour' provisions override the liability that content hosts may face under the Act; and
- The Act does not provide statutory authority for blocking objectionable content online (Department of Internal Affairs, 2020a, pp. 9–11).

The most controversial aspect of the bill is its provision for the Department of Internal Affairs (DIA) to operate a web filter—an electronic system to block public access to 'objectionable' online content. The filter could prevent access to an entire website, or part of a website (Kenny, 2020a),

[10] The repealed s123 of the Crimes Act 1961 included a proviso (s123(3)) that 'it is not an offence against this section to express in good faith and in decent language, or to attempt to establish by arguments used in good faith and conveyed in decent language, any opinion whatever on any religious subject.'

and builds on what DIA has already been doing since 2010 via its Digital Child Exploitation Filtering System (DCEFS) to block websites that host child sexual abuse images (Department of Internal Affairs, 2019). In 2020, DIA awarded the DCEFS contract to Allot Limited, a multinational provider of network intelligence and security solutions, based in Israel (Department of Internal Affairs, 2020b). The new contract includes filtering 'violent extremism content' in addition to the current material (Kenny, 2020b).

Details about the proposed expanded filter are vague. Regulation of the filter's design and operation will sit with DIA. Internet service providers, technical experts, online content hosts and the public will have a say in it, and there will be a review and appeal process (Kenny, 2020a). InternetNZ chief executive Jordan Carter does not have an issue with filters, but with the state mandating one, which he maintains is in tension with the Christchurch Call's commitment to a free, open and secure internet (quoted in Kenny, 2020a). Neither is it clear that operation of the filter will be subject to independent oversight (Rolinson, 2020). The burden of justification lies with the state, to show how the filter will not override fundamental human rights, including freedom of peaceful assembly and association, and freedom of expression (O'Brien & Micek, 2020).[11]

Questions have also been raised about whether a content filter might provoke a public backlash and fuel desire for access, the risk of over-blocking that silences minority voices who already struggle to be heard, and the effectiveness of filters and how this might be measured and evaluated. Kenny (2020a) quotes Marcin Betkier, a law lecturer at Victoria University of Wellington, as conceding that the days of a free and friendly internet have been and gone:

> It's evolved into a structure which is non-transparent for individuals and governments, in which power is held by big companies that mediate interactions between people. They have monetary incentives that don't necessarily support individual or societal wellbeing. And we can't rely on them to self-regulate. We have to draw a line. We can't sit here and say, 'it's impossible, so let's do nothing'.

Chief Censor David Shanks has similarly commented:

> The core of the debate to come is: 'Is the perfect the enemy of the good, in this space?' It's quite clear that no regulatory response is going to be perfect. But does that mean we just give up and do nothing? Or do we think through practical steps that can make things better—if not perfect (quoted in Donovan, 2020).

On March 11, 2021, the Brainbox Institute released a substantial report funded by the New Zealand Law Foundation on a 'better rules' approach and 'legislation as code'[12] (Barraclough, Fraser & Barnes, 2021). The report discusses the Films, Videos, and Publications Classification (Urgent Interim

[11] In considering freedom of peaceful assembly and association, the UN Human Rights Council has declared that the same rights that people have offline must also be protected online (UN Human Rights Council, 2013; 2015; 2018; 2019).

[12] Barraclough, Fraser, & Barnes (2021) define 'better rules' as the application of 'service design' techniques to policy development, including skillsets from computer science and business process modelling, concept modelling, decision flow diagrams and rule statements. By 'law as code' they mean 'machine-consumable legislation' that embodies (interpretations of) the law in code, for example, writing new laws using a process that could generate a computer-implementable output.

Classification of Publications and Prevention of Online Harm) Amendment Bill on pp. 99–104 (paragraphs 419–435), noting that:

- The Bill reflects a trend across a number of jurisdictions that aims to expand the scope of what kinds of information may not be published or accessed on the internet;[13]
- The Bill 'facilitates' setting up a government-backed (either mandatory or voluntary) web filter 'if one is desired in the future', without making the case that such a framework is needed now;
- The filter will be a self-executing system acting with legal force, with legal consequences;
- While there is a public interest in preventing the intentional spread of objectionable material, a national internet filter has serious human and civil rights implications for privacy and freedom of expression; and
- The Bill proposes extremely loose parameters for how the system would operate and what it would apply to.

The authors note that the Bill delegates to secondary legislation (regulation) all the specifics for how the web filter will work, including mechanisms of review and appeal. They argue that a legislative proposal to implement a digital censorship system should have enough detail so it can be scrutinised by members of the public and Parliament before it is enacted. Any Act should impose greater control on how Executive government designs, implements and audits the system, and provide for effective mechanisms of appeal:

> We urge extreme caution in the progress of this Bill through the House and advise that the electronic system of filtering web access be developed in close consultation with non-government actors using a better rules approach before it is enacted as legislation (Barraclough, Fraser, & Barnes, 2021, p. 104, para. 435).

Recommendations of the Royal Commission of Inquiry

The report of the Royal Commission of Inquiry (the Commission) notes limitations of the provisions in the Summary Offences Act:

> Apart from assaults, it applies only to conduct that occurs in a public place. As well, penalties for offences under the Summary Offences Act are low (for example, the maximum penalty for a conviction of offensive behaviour or language is a fine of $1000). Where the maximum penalty is a fine, taking a hate motivation into account during sentencing would not have much practical effect (Royal Commission of Inquiry, 2020a, p. 703).

Because the definition of a public place encompasses only physical locations, a charge of offensive language cannot be brought under the Summary Offences Act against a person who posts material online—even where the post is clearly directed at another individual or group and is visible to other people who are online (Royal Commission of Inquiry, 2020a, p. 35).

[13] Barraclough, Fraser, & Barnes (2021, p. 100, para. 423) reference, for example, legislation described in the Online Harms White Paper in the UK, the EU's proposed Digital Services Act, and Australian legislation criminalising the sharing of 'abhorrent violent material' (among other things).

The Commission further noted (2020a, p. 703) that a hate motivation for offences is not recorded in charges and convictions, even if it has been taken into account during sentencing:

> This means that recorded convictions do not capture the full blameworthiness (culpability) of the offenders. This limits the signalling effect of prosecution and conviction and means possible needs for rehabilitative interventions are not highlighted.

The Commission recommended creating a separate category of 'hate crime' offences, in which a hate motivation is recognised as an element of existing offences in the Summary Offences Act and the Crimes Act. The Commission considered this would signal that such offences are taken seriously and encourage increased reporting of hate offences to New Zealand Police.

In summary, relevant recommendations of the Royal Commission of Inquiry were:

> **39. Amend legislation** to create hate-motivated offences in:
> a) the Summary Offences Act 1981 that correspond with the existing offences of offensive behaviour or language, assault, wilful damage and intimidation; and
> b) the Crimes Act 1961 that correspond with the existing offences of assaults, arson and intentional damage.
>
> **40. Repeal** section 131 of the Human Rights Act 1993 and insert a provision in the Crimes Act 1961 for an offence of inciting racial or religious disharmony, based on an intent to stir up, maintain or normalise hatred, through threatening, abusive or insulting communications with protected characteristics that include religious affiliation.
>
> **41. Amend** the definition of 'objectionable' in section 3 of the Films, Videos, and Publications Classification Act 1993 to include racial superiority, racial hatred and racial discrimination.
>
> **42. Direct** New Zealand Police to revise the ways in which they record complaints of criminal conduct to capture systematically hate-motivations for offending and train frontline staff in:
> a) identifying bias indicators so that they can identify potential hate crimes when they perceive that an offence is hate-motivated;
> b) exploring perceptions of victims and witnesses so that they are in a position to record where an offence is perceived to be hate-motivated; and
> c) recording such hate-motivations in a way which facilitates the later use of section 9(1)(h) of the Sentencing Act 2002.

The Commission's proposed draft wording of a new provision in the Crimes Act 1961 (to replace section 131 of the Human Rights Act 1993) is:

> **Inciting racial or religious disharmony**
> Every person commits an offence and is liable on conviction to imprisonment for a term not exceeding three years who:
> (a) with intent to stir up, maintain or normalise hatred against any group of persons in New Zealand on the ground of the colour, race, or ethnic or national origins or religion of that group of persons; and
> (b) says or otherwise publishes or communicates any words or material that explicitly or implicitly calls for violence against or is otherwise, threatening, abusive, or insulting to such group of persons.

The Commission recommended that religion be included as a 'protected characteristic' without any qualification to protect freedom of opinion and expression along the lines of section 29J of the

United Kingdom's Public Order Act 1986 (Royal Commission of Inquiry, 2020a, pp. 709–710).[14] The Commission commented (p. 710):

> This section has made prosecution for the offence for stirring up religious hatred practically impossible. For this reason we do not support the introduction of an equivalent provision to New Zealand law. We consider that concerns about freedom of expression are met with a high threshold for liability, requiring the prosecution to establish an intention to stir up, maintain or normalise hatred towards members of the protected group and specifically criminalising explicit and implicit calls for violence against such a group.

If the Royal Commission of Inquiry's recommendations are accepted, much will depend on the interpretation of intent to 'maintain or normalise hatred' as opposed to an intent to stir it up (or incite it), and of 'abusive or insulting' as opposed to 'threatening' communications. The recommendation that religion become a 'protected characteristic' *without qualification* in New Zealand's regulation of harmful communication risks departing from an established consensus in international human rights law.

Regulation of harmful communication in other selected jurisdictions

Peter Thompson (2019) notes that political momentum towards state regulation of social media and other digital intermediaries was well under way before the Christchurch terrorist attack and the Christchurch Call. In what follows, I build on and update his summary of some key issues and developments in selected jurisdictions. These include issues of platform monopolies and data protection, and regulatory responses of various sorts to harmful communication.

United Kingdom

In the United Kingdom, the 2017 House of Common Home Affairs Committee report on online hate speech and extremism noted the inapplicability of traditional frameworks of media regulation to social media and digital platform operators and identified a need for more robust measures to police content, because 'the interpretation and implementation of the community standards in practice is too often slow and haphazard' (House of Commons Home Affairs Committee, 2017, p.14).

The report states (p. 24, para 15):

> It is essential that the principles of free speech and open public debate in democracy are maintained—but protecting democracy also means ensuring that some voices are not drowned out by harassment and persecution, by the promotion of violence against particular groups, or by terrorism and extremism.

In 2018, the UK Government also announced the introduction from April 2020 of a digital services tax—a two per cent levy on revenues made by large businesses that provide a social media service, search engine or online marketplace to UK-based users. A large business is defined as generating in-scope annual global revenues of more than £500m, more than £25m of which are attributable to UK sales (Norton Rose Fulbright, 2020).

[14] Similar qualifications are provided in section 18(D) of Australia's Racial Hatred Act 1995, and in section 319(3) of Canada's Criminal Code, which is discussed below (p. 23).

Thompson (2019, p. 86) notes that while the digital services tax was intended primarily to force global tech companies to pay tax that they avoid by declaring profits in offshore havens, it is important because it reclaims online commercial turnover as domestic economic activity.

In April 2019, an Online Harms White Paper was jointly released by the Home Office and the Department for Digital, Culture, Media and Sport (HM Government, 2019). The paper proposed that social media companies and tech firms be legally required to protect their users and face tough penalties if they do not comply. A new independent regulator will be introduced to ensure companies meet their responsibilities.

The White Paper tackles a range of harms, including inciting violence and violent content, encouraging suicide, disinformation, cyber bullying and children accessing inappropriate material. There will be stringent requirements for companies to take even tougher action against terrorist and child sexual exploitation and abuse content. Digital Secretary Jeremy Wright said:

> The era of self-regulation for online companies is over. Voluntary actions from industry to tackle online harms have not been applied consistently or gone far enough. Tech can be an incredible force for good and we want the sector to be part of the solution in protecting their users. However those that fail to do this will face tough action (Department for Digital, Culture, Media and Sport & Home Office, 2019).

The Government's response to consultation on the White Paper was released in December 2020 (UK Government, 2020). Social media companies will need to remove and limit the spread of harmful content. Proposed penalties for not doing so include fines of billions of pounds or being blocked in the UK. Platforms will have to publish an audit of efforts to tackle posts that are harmful but not illegal and abide by a new code of conduct that sets out their responsibilities towards children. The bill requires the most popular sites to set their own terms and conditions, and face fines if they fail to stick to them. Ofcom, a government regulator, will have the power to levy unprecedented fines of up to £18m or 10% of global turnover. And the Government has said it will introduce secondary legislation with criminal sanctions for senior managers, if its desired changes do not eventuate (Hern, 2020; Kelion, 2020).

Digital Secretary Oliver Dowden has undertaken to introduce the bill in 2021.

Australia

In Australia, the Government moved swiftly following the Christchurch mosque shootings to introduce a Criminal Code Amendment (Unlawful Showing of Abhorrent Violent Material) Bill (Attorney-General for Australia, 2019; Chatwood, 2019). The bill proposed two new sets of offences.

- It will be a criminal offence for social media platforms not to remove abhorrent violent material expeditiously. This will be punishable by three years' imprisonment or fines of up to AUD$2.1m for individuals, or the greater of AUD$10.5m and 10% of annual turnover for companies.
- Platforms anywhere in the world must notify the Australian Federal Police if they detect or are made aware that their service is streaming abhorrent violent conduct that is happening in Australia. A failure to do this will be punishable by fines of up to AUD$168,000 for an individual or $840,000 for a corporation.

An eSafety Commissioner has new powers to issue a notice to a content service provider or hosting service stating that, at the time of the notice, the 'abhorrent violent material' could be accessed or was hosted on their service. The provisions apply whether the content service or hosting service is located within or outside Australia.

The bill was enacted on April 4, 2019, assented to on April 5, and came into effect on April 6 as the Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019. Although the Act targets a narrow range of particularly extreme content, the Digital Industry Group Inc. (representing Facebook, Google and Twitter) argued that the bill was rushed through with insufficient deliberation and that United States law prevents the Group from sharing content data (Bogle, 2019; Thompson, 2019, p. 88).

A report three months later by the Australian Competition & Consumer Commission on its digital platforms inquiry (ACCC, 2019) recommended a wider range of measures to address the market power of digital intermediaries. The report had been initiated in December 2017 and recommended:

- Changes to merger law and advance notice of acquisitions;
- Changes to internet search engine and browser defaults;
- Further work on data portability for digital platforms;
- Proactive investigation, monitoring and enforcement of issues in markets in which digital platforms operate;
- An inquiry into the supply of ad tech services and advertising agencies;
- A harmonised media regulatory framework to address regulatory imbalance between news media businesses and digital platforms;
- A requirement that designated digital platforms provide codes of conduct governing relationships between digital platforms and media businesses to the Australian Communications and Media Authority (ACMA);
- A mandatory ACMA take-down code to assist copyright enforcement on digital platforms;
- Stable and adequate funding for public broadcasters, grants for local journalism and tax settings to encourage philanthropic support for journalism;
- Improving digital media literacy in the community, including digital media literacy in schools, monitoring efforts of digital platforms to implement credibility signalling and a Digital Platforms Code to counter disinformation;
- Strengthened protections in the Privacy Act, broader reform of Australian privacy law, an Office of the Australian Information Commissioner privacy code for digital platforms, statutory tort for serious invasions of privacy, prohibition against unfair contract terms and unfair trading practices; and
- A requirement that digital platforms comply with internal dispute resolution requirements, and establishment of an ombudsman scheme to resolve complaints and disputes with digital platform providers.

The Government responded to the ACCC report in December 2019 by:

- Investing $26.9 million in a new special unit in the ACCC to monitor and report on the state of competition and consumer protection in digital platform markets, taking enforcement action as necessary, and undertaking inquiries as directed by the Treasurer, starting with the supply of online advertising and ad tech services;

- Commencing a staged process to reform media regulation towards a platform-neutral regulatory framework covering both online and offline delivery of media content to Australian consumers;
- Addressing bargaining power imbalances between digital platforms and news media businesses by asking the ACCC to work with the relevant parties to develop and implement voluntary codes to address these concerns; and
- Conducting a review of the Privacy Act and ensuring privacy settings empower consumers, protect their data and best serve the Australian economy (Prime Minister of Australia, 2019).

On the recent stoush between the Australian Government and Facebook, see Working paper 21/04, **Challenges in regulating online content**, p. 15.

Canada

Canada's Criminal Code, Part VIII (Offences against the person and reputation), prohibits advocating or promoting genocide (s318) and public incitement of hatred (s319) (Government of Canada, 2020).

It is a criminal offence to communicate statements in any public place that incite hatred against any identifiable group, where this incitement is likely to lead to a breach of the peace. In section 319:

- 'Communicating' includes by telephone, broadcasting or other audible or visible means;
- 'Statements' include words spoken or written or recorded electronically or electro-magnetically or otherwise, as well as gestures, signs or other visible representations;
- A 'public place' is a place to which the public has access as of right or by invitation, express or implied; and
- An 'identifiable group' means (as in s318) any section of the public distinguished by colour, race, religion, national or ethnic origin, age, sex, sexual orientation, gender identity or expression, or mental or physical disability (s319(7)).

It is also a criminal offence to communicate statements, other than in private conversation, that wilfully promote hatred against any identifiable group (s319(2)). 'Hatred' is not defined in either section. Proceedings cannot be instituted under section 319(2) without the consent of the Attorney General, and defences are specified for section 319(2) in section 319(3)—no one shall be convicted if:

- The person establishes that the statements communicated were true;
- In good faith, the person expressed or attempted to establish by an argument an opinion on a religious subject or an opinion based on a belief in a religious text;
- The statements were relevant to any subject of public interest, the discussion of which was for the public benefit, and if on reasonable grounds the person believed them to be true; or
- In good faith, the person intended to point out, for the purpose of removal, matters producing or tending to produce feelings of hatred toward an identifiable group in Canada.

Germany

In 2017, Germany enacted a law to improve law enforcement in social networks—the Network Enforcement Act (Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken, 2017),

commonly referred to as the NetzDG. The law came into force on October 1, 2017, with the objective of combating hate crime, criminal fake news and other criminal content on social media platforms.

The banned content was already illegal in the German Criminal Code (Strafgesetzbuch). The new requirement was that commercial social networks with at least two million users must take responsibility for removing it themselves. Providers must establish a transparent procedure for dealing with complaints about illegal content, be subject to reporting and documentation requirements, must check complaints immediately and delete 'obviously illegal' content within 24 hours, and delete any 'illegal content' after checking it and block it immediately or at least within seven days of receipt of a complaint. Complainants and users must be informed immediately of the decisions taken and the justification for these. The deleted content must be stored for at least 10 weeks for evidential purposes.

Violations attract fines of up to five million euros, and social networks must provide a service agent in Germany, both to the authorities and to enable pursuit of civil proceedings. An administrative offence can also be punished if it is not committed in Germany.

Providers of social networks who receive more than 100 complaints about illegal content in a calendar year must create a German-language report on their handling of the complaints every six months in the Federal Gazette and on their own home pages. In June 2019, German authorities fined Facebook €2 million on the grounds that it failed to disclose information about the full number of hate-speech postings reported in the first half of 2019 (Breeden, 2019).

From the outset, the proposed law triggered debate about two core, post-World War II values in Germany: a prohibition on hate speech, and the preservation of civil liberties such as freedom of expression and privacy (Kinstler, 2018; Delcker, 2020). Critics of the law were concerned that its measures could stifle political speech or be used as a model for authoritarian governments to crack down on online dissent.[15] On the other hand, there are those who say the law does not go far enough to unmask people posting hate speech and bring them to justice.

A NetzDG evaluation report approved by the Cabinet was forwarded to the Bundestag and Bundesrat in September 2020 (Bundesregierung, 2020c). The report concludes that the NetzDG's objectives have been achieved to a large extent, with a clear improvement in complaint management and public accountability of social network providers when dealing with illegal content, even if implementation by providers is sometimes unsatisfactory. It has been estimated that in January 2018, 16 per cent of Facebook's content moderators worldwide were located in Germany, which accounts for only 1.5 per cent of global Facebook users (Turner, 2018). The NetzDG evaluation report found no indications of 'over-blocking', that is, deleting too much content, too quickly, to avoid possible fines.

Some refinements to the regulatory framework have subsequently been proposed, prompted by the murder of Kassel District President Walter Lübcke by a neo-Nazi extremist in June 2019, the attacks in the vicinity of the Halle synagogue in October 2019 and the terrorist shootings by a far-right extremist in Hanau in February 2020.

[15] See, for example, the March 2021 report by Access Now (2021) for the #KeepItOn Coalition on government-mandated internet shutdowns in 2020. The report documents at least 155 shutdowns in 29 countries.

A draft law to combat right-wing extremism and hate crime (Gesetz zur Bekämpfung des Rechtsextremismus und der Hasskriminalität) was approved by the Federal Cabinet in February 2020 as part of a package of measures agreed by the Federal Government on October 30, 2019, following the Halle shooting (Lambrecht, 2020a). The law was approved by the Federal Council (Bundesrat) in July 2020. Large social media providers are now obliged to report death threats, seditious remarks and other criminal content, not merely block or delete it, so potential attackers can be apprehended before they turn to violence. Anti-Semitic motives will be regarded as an aggravating factor. Insults, defamation and slander against local politicians will be punished more severely, and extra protection given to emergency services workers (Bundesregierung, 2020b).

In April 2020, the Federal Government passed a draft law amending the NetzDG (Bundesregierung, 2020a) that:

- Requires more user-friendly channels for reporting complaints about illegal content;
- Supplements obligations for six monthly transparency reports, including information on the use and results of automated procedures for finding and deleting illegal content and access by independent research institutions to anonymised data;
- Increases protection against unauthorised deletion of social media content; and
- Amends s.14 of the Telemedia Act to require a direct right to information from service providers for use as evidence if anyone wants to defend themselves in court against threats or insults (Lambrecht, 2020b).

Like any law, the bill has to be signed by Federal President Frank-Walter Steinmeier. Usually, this is a formality, but Steinmeier is reported to have concerns that Justice Minister Lambrecht's plan to force platforms to hand over sensitive user data to authorities could be deemed unconstitutional by the country's top judges. To date, he has declined to sign the bill (Delcker, 2020). A 'repair bill' is entangled in debates over data disclosure (including URLs) and privacy concerns, led by the Greens and FDP, to the frustration of the SPD and CDU/CSU (Bubrowski, 2021).

Delcker (2020) reflects that it is perhaps no coincidence that debate over what can and cannot be said on the internet is happening in Germany:

> The country has some of the strictest laws on what is acceptable speech. Forged in the late 1950s, Germany's robust hate speech laws were a direct response to the country's experience with Nazism and an acknowledgment that the rise of authoritarianism was partly made possible by the fact that it was legal to use incendiary propaganda that drew on racist tropes and was designed to stoke prejudice.

Yet on the other hand, the need to protect people's right to privacy and civil liberties also resonates strongly in Germany, given the history of surveillance, spying and informing on neighbours both in Nazi Germany and in the Stasi-era of the DDR. Those who are opposed to the NetzDG particularly object to content moderators employed by tech companies making the first call on whether content falls foul of German hate speech laws, which is something judges often struggle to determine, and then being expected to report the authors of the posts to investigators at Germany's Federal Criminal Police Office (Delcker, 2020; Stockmann, 2020, p. 256).

Press freedom advocates have also warned that the NetzDG provides a signal to authoritarian regimes to crack down hard on illegal online content, and a template for censoring political opposition. Two weeks after Germany passed the NetzDG in 2017, Russia approved its own hate speech bill, which referred explicitly to the German law. Since then, more than a dozen countries,

from Venezuela to the Philippines and Malaysia, have passed similar legislation. In August 2020, Turkey passed what the non-profit Electronic Frontier Foundation called 'the worst version of Germany's NetzDG yet' (quoted in Delcker, 2020).

The occupation of the Capitol in Washington DC on January 6, 2021 has, however, given fresh impetus to progressing legislation to combat harmful communication online. Revisions to the legislation are to be debated by the Bundestag and could be passed in 2021 for consideration by the upper house, the Bundesrat (Silk & Connor, 2021). The Bundestag is also amending the Act against Restraints of Competition, to strengthen the Federal Cartel Office (Schwenn, 2021).

France

In France, an interim report setting out a framework for social media regulation was published in May 2019. The report seeks to manage the tension between, on the one hand, freedom of expression and the opportunities social networks allow for individuals and civil society to bypass conventional media and communicate directly and publicly and, on the other hand, unacceptable abuses of those same freedoms.

The regulatory team had been granted six months' partial access to Facebook and concluded:

> Even if the abuses are committed by users, social networks' role in the presentation and selective promotion of content, the inadequacy of their moderation systems and the lack of transparency of their platforms' operation justify intervention by the public authorities, notwithstanding the efforts made by certain operators (Office of the Secretary of State for Digital Affairs, 2019, p. 10).

The report concluded:

> Public intervention to force the biggest players to assume a more responsible and protective attitude to our social cohesion therefore appears legitimate. Given the civil liberty issues at stake, this intervention should be subject to particular precautions. It must (1) respect the wide range of social network models, which are particularly diverse, (2) impose a principle of transparency and systematic inclusion of civil society, (3) aim for a minimum level of intervention in accordance with the principles of necessity and proportionality and (4) refer to the courts for the characterisation of the lawfulness of individual content (p. 2).

The report argued that the public policy response must strike a balance between recognising what platforms are already doing, avoiding a punitive approach, and sending a strong political signal and making social networks more accountable. It therefore proposed a five pillars approach:

- A public regulatory policy guaranteeing individual freedoms and platforms' entrepreneurial freedom;
- A prescriptive regulation focusing on the accountability of social networks, implemented by an independent administrative authority;
- Informed political dialogue between the operators, the government, the legislature and civil society;
- An independent administrative authority, acting in partnership with other branches of the state, and open to civil society; and
- European co-operation, reinforcing member states' capacity to take action on global platforms and reducing the political risks related to implementation in each member state (p. 3).

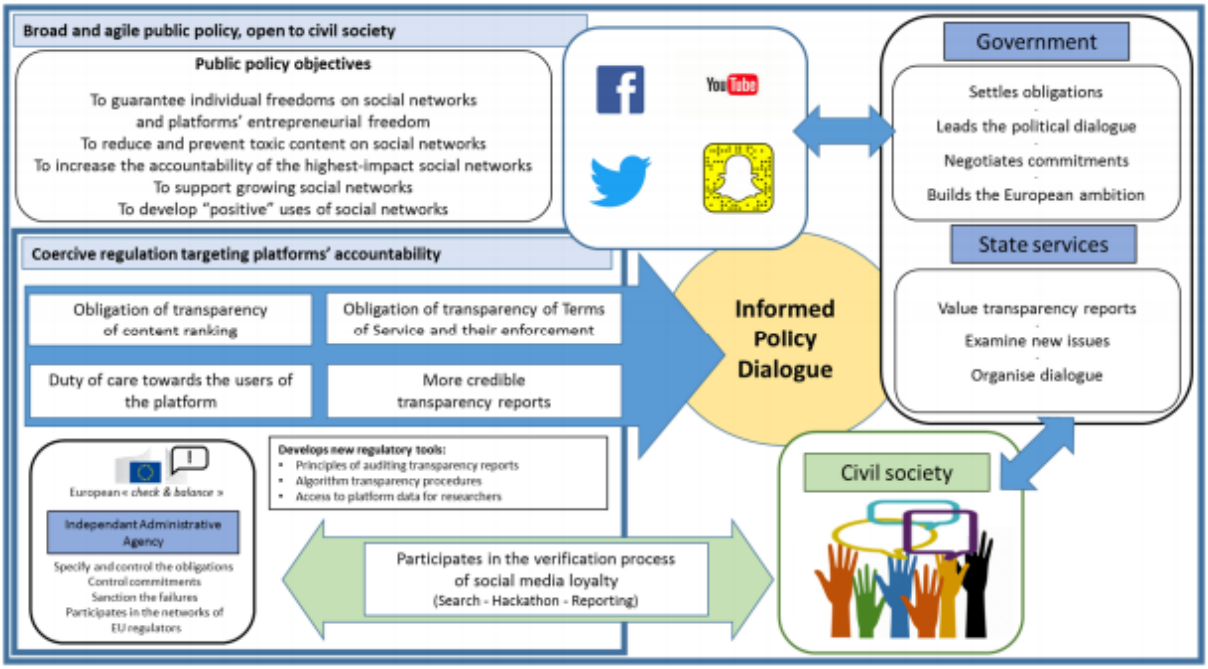

Figure 1: Proposed regulatory framework for social media regulation (Office of the Secretary of State for Digital Affairs, 2019, p. 3).

In July 2019, the lower house of the French Parliament approved an internet regulation bill that would require companies to remove any content that incites or encourages hateful violence or discrimination based on one's race or religion. The bill (Assemblée nationale, 2019) was partly inspired by the German NetzDG. As in Germany, it attracted criticism that by putting the onus for regulating content onto online platforms, and to avoid fines, tech companies may be overzealous in their moderation and harm freedom of expression (Breeden, 2019).

While the Parliament passed the bill in May 2020, it suffered a setback the following month when France's Constitutional Council struck down some of its provisions because they disproportionately infringe on freedom of speech. The flagship provision in the law created an obligation for online platforms to take down hateful content flagged by users within 24 hours. If they failed to do so, they risked fines of up to €1.25 million. The court noted in its ruling that this put the onus for analysing content solely on tech platforms without the involvement of a judge, within a very short time frame, and with the threat of hefty penalties. The court said this created an incentive for risk-averse platforms to indiscriminately remove flagged content, whether or not it was clearly 'hate speech' (Breeden, 2020).

The ruling was an indirect blow to the German NetzDG, which Justice Minister Lambrecht had suggested could serve as a model for Europe-wide rules (Delcker, 2020).

In July 2019, the French Government also approved a digital services tax—three per cent on total sales in France for companies with global sales of over €750m and which make more than €25m a year in France (BBC News, 2019).

European Union

In 2016, the European Commission took a co-regulation approach in its *Code of Conduct on Countering Illegal Hate Speech Online* (European Commission, n.d., 2). The Code of Conduct, like the Christchurch Call, is voluntary and non-binding, and designed to combat the spread of illegal 'hate speech' online in Europe.

Signatories commit to reviewing notifications of illegal 'hate speech' on their platforms and removing or disabling access to this content within 24 hours, educating and raising awareness with users about content that is not permitted under their rules and community guidelines, and sharing best practice with one another.

Facebook, YouTube, Twitter, Microsoft, Instagram and jeuxvideo.com signed up. The fifth evaluation report in 2020 noted that while companies still need to improve their feedback to users' notifications, on average 90 per cent of notifications are reviewed within 24 hours and 71 per cent of the content is removed (European Commission, 2020a). The Code of Conduct has enabled some common approaches in the drafting of rules across the various platforms, but there has been little consistency in enforcement (Flew, Martin & Suzor, 2019, pp. 40–41).

As part of the European Digital Strategy, *Shaping Europe's Digital Future*, the European Commission announced in December 2020 that it would upgrade the rules governing digital services in the European Union. Two legislative initiatives were proposed—a Digital Services Act and a Digital Markets Act (European Commission, 2020c; Riegert, 2020; Satariano, 2020). These build on the EU's General Data Protection Regulation, which was adopted in May 2016 and came into effect in all member states in May 2018 to harmonise data privacy laws across Europe (European Commission, n.d., 1).

The Digital Services Act and Digital Markets Act have two main goals:

- To create a safer digital space in which the fundamental rights of all users of digital services are protected; and
- To establish a level playing field to foster innovation, growth, and competitiveness, both in the European Single Market and globally.

Under the Digital Services Act, binding EU-wide obligations will apply to all digital services that connect consumers to goods, services or content, including new procedures for faster removal of illegal content as well as comprehensive protection for users' fundamental rights online. The Act will introduce new, harmonised EU-wide obligations, graduated on the basis of those services' size and impact, such as:

- Rules for the removal of illegal goods, services or content online;
- Safeguards for users whose content has been erroneously deleted by platforms;
- New obligations for very large platforms to take risk-based action to prevent abuse of their systems;
- Wide-ranging transparency measures, including on online advertising and on the algorithms used to recommend content to users;
- New powers to scrutinise how platforms work, including by facilitating access by researchers to key platform data;

- New rules on traceability of business users in online market places, to help track down sellers of illegal goods or services; and
- A co-operation process among public authorities to ensure effective enforcement across the EU.

Platforms that reach more than 10% of the EU's population (45 million users) are to be subject not only to specific obligations to control their own risks, but also to a new oversight structure—a board of national Digital Services Co-ordinators, with special powers for the Commission in supervising very large platforms, including the ability to sanction them directly.

The Digital Markets Act will harmonise rules defining and prohibiting uncompetitive practices and provide an enforcement mechanism based on market investigations. It will:

- Apply only to major providers of the core platform services most prone to unfair practices, such as search engines, social networks or online intermediation services, which meet objective legislative criteria to be designated as gatekeepers;
- Define quantitative thresholds as a basis to identify presumed gatekeepers—and the Commission will have powers to designate companies as gatekeepers following a market investigation;
- Prohibit a number of unfair practices, such as blocking users from un-installing any pre-installed software or apps;
- Require gatekeepers to proactively put in place certain measures, such as targeted measures allowing the software of third parties to properly function and interoperate with their own services;
- Impose sanctions for non-compliance, which could include fines of up to 10% of the gatekeeper's worldwide turnover, to ensure the effectiveness of the new rules—for recurrent infringers, these sanctions may also involve the obligation to take structural measures, potentially extending to divestiture of certain businesses, where no other equally effective alternative measure is available to ensure compliance; and
- Allow the Commission to carry out targeted market investigations to assess whether new gatekeeper practices and services need to be added to these rules, to ensure that the new gatekeeper rules keep up with the fast pace of digital markets (European Commission, 2020c).

The process to legislate these proposals is likely to be protracted and fraught, with tech companies lobbying to resist change (Riegert, 2020; Schwenn, 2021). Debate on them will also test the relationship between the European Union and the United States, which was strained during the Trump administration on issues like digital taxes (Satariano, 2020).[16]

[16] The Trump administration reacted negatively to digital services tax regulations because most of the digital intermediaries falling under the new regimes are based in the United States. President Trump ordered an investigation into France's planned digital services tax—a move that could result in retaliatory tariffs (BBC News, 2019).

On December 8, 2020, responsible Ministers of all EU member states signed the Berlin Declaration on Digital Society and Value-based Digital Government (European Commission, 2020b). This asserts a role for the public sector in driving a values-based digital transformation of European societies. Seven key principles are:

- **Validity and respect** of fundamental rights and democratic values in the digital sphere;
- **Social participation and digital inclusion** to shape the digital world;
- **Empowerment and digital literacy**, allowing all citizens to participate in the digital sphere;
- **Trust and security in digital government interactions**, allowing everyone to navigate the digital world safely, authenticate and be digitally recognised within the EU conveniently;
- **Digital sovereignty and interoperability**, as a key in ensuring the ability of citizens and public administrations to make decisions and act self-determined in the digital world;
- **Human-centred systems and innovative technologies in the public sector**, strengthening its pioneering role in the research on secure and trustworthy technology design; and
- **A resilient and sustainable digital society**, preserving our natural foundations of life in line with the Green Deal and using digital technologies to enhance the sustainability of our health systems (European Commission, 2020b).

At the World Economic Forum in January 2021, European Commission president Ursula von der Leyden urged the US to join the EU's attempts to 'contain the immense power' of Big Tech and create a 'digital economy rule book'. She said Brussels wants social media giants to disclose how their business models and algorithms work and criticised Twitter's decision to remove Donald Trump's account as a 'serious interference with freedom of expression', asserting that regulators should create a 'framework of laws for such far-reaching decisions' (Financial Times reporters, 2021).

UN Secretary-General António Guterres has expressed similar views. He launched a *Roadmap for digital co-operation* in June 2020, and in January 2021, called for a global regulatory framework for Big Tech, so that decisions about de-platforming are made under the rule of law rather than by private companies (Guterres, 2021).

Conclusion: Regulate with restraint

A free, open and democratic society protects everyone's right to freedom of opinion and expression but may justifiably qualify this freedom to prevent harm to others, if it does so in ways that conform to strict tests of legality, proportionality and necessity.

There is an established consensus in international human rights law that it may be justifiable to restrict public communication that incites discrimination, hostility or violence against a social group with a common 'protected characteristic' such as nationality, race or religion.

Regulation to protect people from criticism, offence or lack of respect is not a justifiable restriction of freedom of expression, and international human rights standards specifically discourage blasphemy laws. The recommendation of the Royal Commission of Inquiry that New Zealand's regulation of harmful communication include religion as a 'protected *characteristic' without qualification* risks departing from an established consensus in international human rights law.

International human rights standards urge careful distinctions between (1) communication that is illegal and a criminal offence; (2) communication that is not a criminal offence but that may justify a

civil suit; and (3) so-called 'lawful hate speech' that does not give rise to criminal or civil sanctions, but may still raise concerns about tolerance, civility and respect for others.

Much harmful communication is already a criminal offence under existing law. Before rushing to pass new laws or amend existing ones, it would be wise to pause and consider arguments for and against regulation of harmful communication, and alternatives to prohibition and censorship.

These issues will be explored further in Working paper 21/06, **Arguments for and against restricting freedom of expression**, Working paper 21/07, **Striking a fair balance in regulation of harmful communication**, and the final paper in the series, Working paper 21/08, **Counter-speech and civility as everyone's responsibility**.

References

ACCC. (2019) Digital platforms inquiry: Final report. Australian Competition & Consumer Commission, July 26, 2019. Accessed October 12, 2020, from https://www.accc.gov.au/focus-areas/inquiries-ongoing/digital-platforms-inquiry

Access Now. (2021). *Shattered dreams and lost opportunities: A year in the fight to #KeepItOn*. Accessed March 4, 2021, from https://www.accessnow.org/cms/assets/uploads/2021/03/KeepItOn-report-on-the-2020-data_Mar-2021_1.pdf

Appiah, K. (2021). Is it OK that my sister secretly records our dad for laughs? *New York Times*, January 21, 2021. Accessed January 22, 2021, from https://nyti.ms/2Y6zklT

Article 19. (2009). *The Camden principles on freedom of expression and equality*. London: Article 19. Accessed November 2, 2020, from https://www.article19.org/data/files/pdfs/standards/the-camden-principles-on-freedom-of-expression-and-equality.pdf

Assemblée nationale. (2019). Texte nº 2062, adopté par la commission, sur la proposition de loi de Mme Laetitia Avia et plusieurs de ses collègues visant à lutter contre la haine sur internet (1785). Accessed October 14, 2020, from http://www.assemblee-nationale.fr/dyn/15/textes/l15b2062_texte-adopte-commission

Aswad, E. (2013). To ban or not to ban blasphemous videos. *Georgetown Journal of International Law, 44*(4), 1313–1328.

Attorney-General for Australia. (2019). Tough new laws to protect Australians from live-streaming violent crimes. Media release, March 30, 2019. Accessed October 14, 2020, from https://www.attorneygeneral.gov.au/media/media-releases/tough-new-laws-protect-australians-live-streaming-violent-crimes-3-march-2019

Barraclough, T., Fraser, H., & Barnes, C. (2021). *Legislation as code for New Zealand: Opportunities, risks, and recommendations.* Brainbox Institute. Accessed March 13, 2021, from https://www.brainbox.institute/s/Legislation-as-Code-9-March-2021-for-distribution.pdf

BBC News. (2019). France tech tax: What's being done to make internet giants pay more? July 11, 2019. Accessed October 14, 2020, from https://www.bbc.com/news/business-48928782

Black, J. (2020). What really happened in the Sir Bob Jones v Renae Maihi defamation case? *New Zealand Listener,* February 29, 2020. Archived from https://www.noted.co.nz/currently/currently-social-issues/sir-bob-jones-what-really-happened-in-renae-maihi-case

Bogle, A. (2019). Laws targeting terror videos on Facebook and YouTube 'rushed' and 'knee-jerk', lawyers and tech industry say. *ABC News,* April 3, 2019. Accessed October 14, 2020, from https://www.abc.net.au/news/science/2019-04-04/facebook-youtube-social-media-laws-rushed-and-flawed-critics-say/10965812

Breeden, A. (2019). France will debate a bill to stop online hate speech: What's at stake? *New York Times*, July 1, 2019. Accessed October 14, 2020, from https://nyti.ms/2KR4n05

Breeden, A. (2020). French court strikes down most of online hate speech law. *New York Times*, June 18, 2020. Accessed October 14, 2020, from https://nyti.ms/3dlBeUV

Bubrowski, H. (2021). Hasskriminalität: Viel Hass, kein Gesetz. *Frankfurter Allgemeine Zeitung*, February 12, 2021. Accessed February 15, 2021, from https://www.faz.net/-gpg-a8jto

Bundesregierung. (2020a). Gezielte Bekämpfung von Hasskriminalität. April 1, 2020. Accessed October 11, 2020, from https://www.bundesregierung.de/breg-de/aktuelles/bekaempfung-hasskriminalitaet-1738150

Bundesregierung. (2020b). Gesetzespaket gegen Hasskriminalität: Entschieden gegen Hetze im Netz. July 3, 2020. Accessed October 14, 2020, from https://www.bundesregierung.de/breg-de/aktuelles/gesetz-gegen-hasskriminalitaet-1722896

Bundesregierung. (2020c). Netzwerkdurchsetzungsgesetz: Ziele des Gesetzes wurden erreicht. September 9, 2020. Accessed October 13, 2020, from https://www.bundesregierung.de/breg-de/aktuelles/netzwerkdurchsetzungsgesetz-1783308

Caldwell, J. (2012). Holocaust denial in New Zealand: A study of three recent cases at New Zealand universities. *Holocaust Studies: A journal of culture and history, 18*(1), 119–141. https://doi.org/10.1080/17504902.2012.11087298

Chatwood, R. (2019). Australia suddenly passes new laws regulating streaming of abhorrent violent material by ISPs and other content providers. *Dentons*, April 15, 2019. Accessed October 14, 2020, from https://www.dentons.com/en/insights/alerts/2019/april/15/australia-suddenly-passes-new-laws-regulating-streaming-of-abhorrent-violent-material

Delcker, J. (2020). Germany's balancing act: Fighting online hate while protecting free speech. *Politico,* October 1, 2020; updated October 6, 2020. Accessed October 13, 2020, from https://www.politico.eu/article/germany-hate-speech-internet-netzdg-controversial-legislation/?fbclid=IwAR3MKBBpyNrO0hiqzJGULPjeYqDV9O9%E2%80%A6

Department for Digital, Culture, Media and Sport & Home Office. (2019). UK to introduce world first online safety laws. Media release April 8, 2019. Accessed October 14, 2020, from https://www.gov.uk/government/news/uk-to-introduce-world-first-online-safety-laws

Department of Internal Affairs (2019). Internet and website filter. Accessed October 16, 2020, from https://www.dia.govt.nz/Censorship-DCEFS

Department of Internal Affairs. (2020a). Regulatory impact statement: Countering violent extremism online—changes to censorship legislation to better protect New Zealanders from online harm. January 6, 2020. Accessed October 16, 2020, from https://www.dia.govt.nz/diawebsite.nsf/Files/Proactive-releases/$file/regulatory-impact-assessment-countering-violent-extremism-online.pdf

Department of Internal Affairs. (2020b). New service provider for Digital Child Exploitation Filter. Media release, August 5, 2020. Accessed October 16, 2020, from https://www.dia.govt.nz/press.nsf/d77da9b523f12931cc256ac5000d19b6/eb783be28ff48661cc2585bb007bc060!OpenDocument

Dillane, T. (2021). Swastika on milk worker's overalls after Christchurch attack leads to sacking. *NZ Herald,* January 23, 2021. Accessed January 25, 2021, from https://www.nzherald.co.nz/nz/swastika-on-milk-workers-overalls-after-christchurch-attack-leads-to-sacking/4IQ36E4WYFEYOLLH3BM4MCHYKY/

Donovan, E. (2020). The Detail: Concern over New Zealand's new internet censorship laws. *Stuff,* June 10, 2020. Accessed October 16, 2020, from https://www.stuff.co.nz/national/the-detail/300031018/the-detail-concern-over-new-zealands-new-internet-censorship-laws

Duff, M. (2019). Hate crime law review fast-tracked following Christchurch mosque shootings. *Stuff*, March 30, 2019. Accessed October 15, 2020, from https://www.stuff.co.nz/national/christchurch-shooting/111661809/hate-crime-law-review-fasttracked-following-christchurch-mosque-shootings

Edgeler, G. (2020). On the possibility of laws further regulating hate speech. *Public Address,* February 17, 2020. Accessed October 15, 2020, from https://publicaddress.net/legalbeagle/on-the-possibility-of-laws-further-regulating/

European Commission. (n.d., 1). Data protection in the EU. Accessed January 19, 2021, from https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu_en

European Commission. (n.d., 2). The EU code of conduct on countering illegal hate speech online. Accessed January 20, 2021, from https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en

European Commission. (2020a). Countering illegal hate speech online: 5th evaluation of the Code of Conduct. June 2020. Accessed October 15, 2020, from https://ec.europa.eu/info/sites/info/files/codeofconduct_2020_factsheet_12.pdf

European Commission. (2020b). Berlin Declaration on Digital Society and Value-based Digital Government, December 11, 2020. Accessed January 19, 2021, from https://ec.europa.eu/digital-single-market/en/news/berlin-declaration-digital-society-and-value-based-digital-government

European Commission. (2020c). The Digital Services Act package. European Commission, December 16, 2020. Accessed January 18, 2020, from https://ec.europa.eu/digital-single-market/en/digital-services-act-package

European Union. (2008). Council Framework Decision 2008/913/JHA of 28 November 2008 on combating certain forms and expressions of racism and xenophobia by means of criminal law. Accessed January 18, 2021, from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32008F0913

Feinberg, J. (1973). *Social philosophy*. Englewood Cliffs, NJ: Prentice-Hall.

Financial Times reporters. (2021). Davos highlights: European leaders urge Biden to extend efforts to reignite international co-operation. *Financial Times,* January 26, 2021. Accessed January 28, 2021, from https://www.ft.com/content/02465195-1957-490d-a3c8-4c54d45469a9

Flew, T., Martin, F., & Suzor, N. (2019). Internet regulation as media policy: Rethinking the question of digital communication platform governance. *Journal of Digital Media & Policy, 10*(1), 33–50. https://doi.org/10.1386/jdmp.10.1.33_1

Fraser, David. (2009). 'On the internet, nobody knows you're a Nazi': Some comparative legal aspects of holocaust denial on the WWW. In I. Hare & J. Weinstein (Eds.), *Extreme speech and democracy* (pp. 511–537). Oxford: Oxford University Press.

Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken (2017). Accessed October 13, 2020, from https://www.gesetze-im-internet.de/netzdg/BJNR335210017.html

Government of Canada. (2020). Criminal Code. Revised Statutes of Canada, 1985. Justice laws website, October 29, 2020. Accessed January 18, 2020, from https://laws.justice.gc.ca/eng/acts/C-46/page-68.html

Guterres, A. (2021). Secretary-General's press conference on his priorities for 2021, January 28, 2021. Accessed February 1, 2021, from https://www.un.org/sg/en/content/sg/press-encounter/2021-01-28/secretary-generals-press-conference-his-priorities-for-2021

Hansen v R. (2007). NZSC 7.

Hern, A. (2020). Online harms bill: firms may face multibillion-pound fines for illegal content. *The Guardian,* December 15, 2020. Accessed January 21, 2020, from https://www.theguardian.com/technology/2020/dec/15/online-harms-bill-firms-may-face-multibillion-pound-fines-for-content

Herz, M., & Molnar, P. (2012). Introduction. In M. Herz & P. Molnar (Eds), *The content and context of hate speech: Rethinking regulation and responses* (pp. 1–7). Cambridge: Cambridge University Press.

HM Government. (2019). *Online harms white paper*. Secretary of State for Digital, Culture, Media & Sport and Secretary of State for the Home Department, April 2019. Accessed October 14, 2020, from https://www.gov.uk/government/consultations/online-harms-white-paper

House of Commons Home Affairs Committee. (2017). Hate crime: abuse, hate and extremism online. Fourteenth report of session 2016–17, April 25, 2017. Accessed October 12, 2020, from https://publications.parliament.uk/pa/cm201617/cmselect/cmhaff/609/609.pdf

Human Rights Commission. (n.d.). Racially offensive comments. Accessed November 2, 2020, from https://www.hrc.co.nz/enquiries-and-complaints/faqs/racially-offensive-comments/

Human Rights Commission. (2019). *Whakamauāhara Hate Speech. An overview of the current legal framework*. December 17, 2019. Accessed October 15, 2020, from https://www.hrc.co.nz/news/resource-hate-speech-legal-framework-published/

Kelion, L. (2020). Online harms law to let regulator block apps in UK. *BBC News,* December 15, 2020. Accessed January 18, 2021, from https://www.bbc.com/news/technology-55302431

Kenny, K. (2020a). The Government wants to filter the internet. Is that a good idea? *Stuff,* July 24, 2020. Accessed October 5, 2020, from https://www.stuff.co.nz/technology/122136945/the-government-wants-to-filter-the-internet-is-that-a-good-idea

Kenny, K. (2020b). Concerns about company awarded New Zealand's internet filter contract. *Stuff,* August 22, 2020. Accessed October 16, 2020, from https://www.stuff.co.nz/technology/122524249/concerns-about-company-awarded-new-zealands-internet-filter-contract

Kidd, R. (2021). Man claims racist abuse directed at Indian family was 'light-hearted'. *NZ Herald,* February 11, 2021. Accessed February 12, 2021, from https://www.nzherald.co.nz/nz/man-claims-racist-abuse-directed-at-indian-family-was-light-hearted/FYQMNFC5SN6YYY665PCH4SE64E/

Kinstler, L. (2018). Germany's attempt to fix Facebook is backfiring. *The Atlantic*, May 18, 2018. Accessed October 14, 2020, from https://www.theatlantic.com/international/archive/2018/05/germany-facebook-afd/560435/

Lambrecht, C. (2020a). Gesetzespaket gegen Rechtsextremismus und Hasskriminalität. Bundesministerium der Justiz und für Verbraucherschutz, February 19, 2020. Accessed October 14, 2020, from https://www.bmjv.de/SharedDocs/Artikel/DE/2020/021920_Kabinett_Bekaempfung_Rechtsextremismus_Hasskriminalitaet.html

Lambrecht, C. (2020b). Weiterentwicklung des Netzwerkdurchsetzungsgesetzes. Bundesministerium der Justiz und für Verbraucherschutz, April 1, 2020. Accessed October 14, 2020, from https://www.bmjv.de/SharedDocs/Artikel/DE/2020/040120_NetzDG.html

Living Word Distributors Ltd v Human Rights Action Group. (2000). Wellington, 3 NZLR 570.

Manch, T. (2020). New Zealand twice halts publication of Christchurch mosque shooter's manifesto, in Ukraine and Italy. *Stuff,* September 3, 2020. Accessed October 15, 2020, from https://www.stuff.co.nz/national/politics/122627897/new-zealand-twice-halts-publication-of-christchurch-mosque-shooters--manifesto-in-ukraine-and-italy

Mason, C., & Errington, K. (2019). *Anti-social media: Reducing the spread of harmful content on social media networks*. Auckland: Helen Clark Foundation. Accessed October 15, 2020, from https://helenclark.foundation/reports/anti-social-media/

Molnar, P. (2012). Interview with Kenan Malik. In M. Herz & P. Molnar (Eds), *The content and context of hate speech: Rethinking regulation and responses* (pp. 81–91). Cambridge: Cambridge University Press.

Nagel, T. (2002). *Concealment and exposure, And other essays*. Oxford: Oxford University Press.

Norton Rose Fulbright. (2020). The UK's digital services tax: What's new. May 2020. Accessed October 14, 2020, from https://www.nortonrosefulbright.com/en-de/knowledge/publications/24da19c2/the-uks-digital-services-tax-whats-new

O'Brien, L., & Micek, P. (2020). *Defending peaceful assembly and association in the digital age: Takedowns, shutdowns, and surveillance*. Access Now, July 2020. Accessed October 20, 2020, from https://www.accessnow.org/cms/assets/uploads/2020/07/Defending-Peaceful-Assembly-Association-Digital-Age.pdf

Office of the Secretary of State for Digital Affairs. (2019). *Creating a French framework to make social media platforms more accountable: Acting in France with a European vision*. May 2019. Accessed October 14, 2020, from https://www.numerique.gouv.fr/actualites/remise-du-rapport-de-la-mission-de-regulation-des-reseaux-sociaux/

Parekh, B. (2012). Is there a case for banning hate speech? In M. Herz & P. Molnar (Eds), *The content and context of hate speech: Rethinking regulation and responses* (pp. 37–56). Cambridge: Cambridge University Press.

Parliamentary Services. (2020). Films, Videos and Publications Classification (Urgent Interim Classification of Publications and Prevention of Online Harm) Amendment Bill 2020 268-1. Bills Digest 2626, June 15, 2020. Accessed October 16, 2020, from https://www.parliament.nz/resource/en-NZ/52PLLaw262611/c5893f61c738c39a8ed9693611fcbdf758b21b8e

Prime Minister of Australia. (2019). Response to digital platforms inquiry. Media release, December 12, 2019. Accessed October 14, 2020, from https://www.pm.gov.au/media/response-digital-platforms-inquiry

Riegert, B. (2020). EU takes on tech giants. *Deutsche Welle,* December 16, 2020. Accessed December 16, 2020, from https://www.dw.com/en/opinion-eu-takes-on-tech-giants/a-55965899

Rolinson, C. (2020). Against the great Kiwi firewall: Why I'm concerned about Internal Affairs proposal to filter 'objectionable' internet. *Daily Blog*, June 4, 2020. Accessed October 20, 2020, from https://thedailyblog.co.nz/2020/06/04/against-the-great-kiwi-firewall-why-im-concerned-about-internal-affairs-proposal-to-filter-objectionable-internet/

Royal Commission of Inquiry. (2020a). *Ko tō tātou kāinga tēnei.[This is our home.] Report: Royal Commission of Inquiry into the terrorist attack on Christchurch masjidain on 15 March 2019*. November 26, 2020. Accessed December 8, 2020, from https://christchurchattack.royalcommission.nz/

Royal Commission of Inquiry. (2020b). *Hate speech and hate crime related legislation.* [Companion paper to 2020a]. Accessed December 10, 2020, from https://christchurchattack.royalcommission.nz/publications/comp/introduction/

Satariano, A. (2020). Big fines and strict rules unveiled against 'big tech' in Europe. *New York Times*, December 15, 2020. Accessed December 16, 2020, from https://www.nytimes.com/2020/12/15/technology/big-tech-regulation-europe.html

Schauer, F. (2012). Social epistemology, holocaust denial, and the post-Millian calculus. In M. Herz & P. Molnar (Eds), *The content and context of hate speech: Rethinking regulation and responses* (pp. 129–143). Cambridge: Cambridge University Press.

Schwenn, K. (2021). Scarfes Schwert gegen Digitalbarone. *Frankfurter Allgemeine Zeitung,* January 14, 2021. Accessed January 18, 2021, from https://www.faz.net/-gqe-a7grf

Silk, J., & Connor, R. (2021). Capitol Hill riots prompt Germany to revisit online hate speech law. *Deutsche Welle,* January 8, 2021. Accessed January 18, 2021, from https://www.dw.com/en/capitol-hill-riots-prompt-germany-to-revisit-online-hate-speech-law/a-56171516

Stockmann, D. (2020). Media or corporations? Social media governance between public and commercial rationales. In H. Anheier & T. Baums (Eds), *Advances in corporate governance: Comparative perspectives* (pp. 249–268). Oxford University Press. https://doi.org/10.1093/oso/9780198866367.003.0011

Strossen, N. (2018). *Hate: Why we should resist it with free speech, not censorship*. New York: Oxford University Press.

Suk, J. (2012). Denying experience: Holocaust denial and the free-speech theory of the state. In M. Herz & P. Molnar (Eds), *The content and context of hate speech: Rethinking regulation and responses* (pp. 144–163). Cambridge: Cambridge University Press.

Thompson, P. (2019). Beware of geeks bearing gifts: Assessing the regulatory response to the Christchurch Call. *Political Economy of Communication, 7*(1), 83–104. Accessed October 7, 2020, from https://iamcr.org/s-wg/section/poe/journal-7-1

Turner, Z. (2018). Facebook, Google have a tough new job in Germany: Content cop. *Wall Street Journal*, January 10, 2018. Accessed October 13, 2020, from https://www.wsj.com/articles/facebook-google-have-a-tough-new-job-in-germany-content-cop-1515605207

UK Government. (2020). Consultation outcome: Online Harms White Paper, December 15, 2020. Accessed January 18, 2021, from https://www.gov.uk/government/consultations/online-harms-white-paper

UN Committee on the Elimination of Racial Discrimination. (2013). General recommendation No. 35, Combating racist hate speech, September 26, 2013. Accessed October 30, 2020, from https://undocs.org/CERD/C/GC/35

UN General Assembly. (1948). Universal Declaration of Human Rights. Accessed October 30, 2020, from https://www.un.org/en/universal-declaration-human-rights/

UN General Assembly. (1965). International Convention on the Elimination of all Forms of Racial Discrimination. Accessed November 18, 2020, from https://treaties.un.org/Pages/ViewDetails.aspx?src=TREATY&mtdsg_no=IV-2&chapter=4&clang=_en#14

UN General Assembly. (1966). International Covenant on Civil and Political Rights. Accessed October 30, 2020, from https://treaties.un.org/doc/publication/unts/volume%20999/volume-999-i-14668-english.pdf

UN General Assembly. (2011a). Resolution adopted by the Human Rights Council, 16/18: Combating intolerance, negative stereotyping and stigmatization of, and discrimination, incitement to violence and violence against, persons based on religion or belief. UN Doc A/HRC/RES/16/18, April 12, 2011. Accessed December 17, 2020, from https://undocs.org/en/A/HRC/RES/16/18

UN General Assembly. (2011b). Report of the Special Rapporteur [Frank La Rue] on the promotion and protection of the right to freedom of opinion and expression. UN Doc A/66/290, August 10, 2011. Accessed October 30, 2020, from https://undocs.org/A/66/290

UN General Assembly. (2012a). Report of the Special Rapporteur [Frank La Rue] on the promotion and protection of the right to freedom of opinion and expression. UN Doc A/67/357, September 7, 2012. Accessed November 2, 2020, from https://www.palermo.edu/cele/pdf/SRs-Report.pdf

UN General Assembly. (2012b). Combating intolerance, negative stereotyping, stigmatization, discrimination, incitement to violence and violence against persons, based on religion or belief. Resolution 67/178, December 20, 2012. Accessed December 17, 2020, from https://undocs.org/en/A/RES/67/178

UN General Assembly. (2013). Annual report of the United Nations High Commissioner for Human Rights. Annex: Report of the United Nations High Commissioner for Human Rights on the expert workshops on the prohibition of incitement to national, racial or religious hatred. Appendix: Rabat Plan of Action on the prohibition of advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence. UN Docs A/HRC/22/17/Add.4, January 11, 2013. Accessed November 2, 2020, from https://www.ohchr.org/Documents/Issues/Opinion/SeminarRabat/Rabat_draft_outcome.pdf

UN General Assembly. (2019). Report of the Special Rapporteur [David Kaye] on the promotion and protection of the freedom of opinion and expression. UN Doc A/74/486, October 9, 2019. Accessed October 30, 2020, from https://www.ohchr.org/Documents/Issues/Opinion/A_74_486.pdf

UN Human Rights Committee. (2011). International Covenant on Civil and Political Rights: General comment No. 34. Article 19: Freedom of opinion and expression. UN Doc. CCPR/C/GC/34, September 12, 2011. Accessed October 30, 2020, from https://www2.ohchr.org/english/bodies/hrc/docs/gc34.pdf

UN Human Rights Committee. (2017). *Rabbae v The Netherlands,* Communication No. 2124/2011. UN Doc CCPR/C/117/D/2124/2011, March 29, 2017. Accessed October 30, 2020, from https://juris.ohchr.org/Search/Details/2153

UN Human Rights Council. (2013). Resolution 22/31: Combating intolerance, negative stereotyping and stigmatization of, and discrimination, incitement to violence and violence against, persons based on religion or belief. UN Doc A/HRC/RES/22/31, April 15, 2013. Accessed December 17, 2020, from https://undocs.org/en/A/HRC/RES/22/31

UN Human Rights Council. (2015). Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression. UN Doc. A/HRC/29/32, May 22, 2015. Accessed October 20, 2020, from https://undocs.org/A/HRC/29/32

UN Human Rights Council. (2018). The promotion, protection and enjoyment of human rights on the Internet. UN Doc. A/HRC/38/L.10/Rev.1, July 4, 2018. Accessed October 20, 2020, from https://undocs.org/en/A/HRC/38/L.10/REV.1

UN Human Rights Council. (2019). Report of the Special Rapporteur on the rights to freedom of peaceful assembly and of association. U.N. Doc. A/HRC/41/41, May 17, 2019. Accessed October 20, 2020, from https://undocs.org/A/HRC/41/41

UN OHCHR. (n.d.). Office of the High Commissioner for Human Rights. Status of ratification: Interactive dashboard. Accessed October 30, 2020, from https://indicators.ohchr.org/

UN OHCHR. (2019). Joint open letter on concerns about the global increase in hate speech. [September 23, 2019]. Accessed November 2, 2020 from https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=25036&LangID=E and https://news.un.org/en/story/2019/09/1047102