GEORGE MIRO SHIRTCLIFFE

THE PROPOSED REFORM OF NEW ZEALAND'S CRIMINAL HATE SPEECH LAW

Submitted for the LLB (Honours) Degree

Faculty of Law

Victoria University of Wellington

2022

I Abstract

This paper discusses the Labour Government's 2021 proposal to replace s 131 of the Human Rights Act 1993, New Zealand's current criminal hate speech law, with a new provision in the Crimes Act 1961. The central aim of this paper is to determine whether the provision correctly addresses hate speech. This paper analyses hate speech and whether its harm justifies legislating against it. It agrees the harm it produces is tangible and an increasing issue in New Zealand, indicating s 131 is inadequate. Hate speech laws infringe upon freedom of expression, a justifiably important right protected by the New Zealand Bill of Rights Act 1990. However, s 5 allows for "demonstrably justified" limitations upon it. To achieve this, the proposed provision must be narrowly worded to ensure it is a minimal infringement upon the right. The difficulty in distinguishing insulting communications from hate speech makes this increasingly necessary. Against this normative background, the paper examines the proposal. It agrees that the proposal's placement in the Crimes Act will better signal the unacceptability and real harm of hate speech. Allowing the communication to be made by any means is the most justified change, as it addresses the rapid increase in online hate speech. The proposal's use of "hatred" reflects foreign hate speech legislation and international obligations; however, it must be clearly defined to ensure a narrow application. This paper argues the omission of the "likely to" result element is unjustified, as it removes an important safeguard against an unintentionally wide application. It concludes the proposal has merit but requires more work to develop a narrow offence that presents a demonstrably justified limitation upon freedom of expression.

Key terms: "hate speech", "harm of hate speech", "freedom of expression", "section 131 Human Rights Act 1993"

Table of Contents

I ABSTRACT 2
II INTRODUCTION 4
III THE PROPOSAL 5
A Rationale for Replacing s 131 5
1 Amending s 131 7
B The Proposed Changes 7
1 Means of Communication 8
2 Terminology 8
3 Removal of the "likely to" Result Element 10
C Summary 10
IV HATE SPEECH 11
A Defining Hate Speech 11
B The Harm of Hate Speech 14
C Freedom of Expression 17
1 Democratic Self-Governance 18
2 Marketplace of Ideas 19
3 Harm of Hate Speech as an Infringement on Free Speech 20
D The Need for Caution 21
E Summary 23
V ANALYSIS 23
A Means of Communication 24
B Hatred… 24
C Removal of the Result Element 26
VI CONCLUSION 30
VII BIBLIOGRAPHY 32
A Cases… 32
1 New Zealand 32
2 Australia 32
3 Canada 32
4 United Kingdom 32
5 United States 32
B Legislation 32
1 New Zealand 32
2 Canada 32
3 Germany 33
4 United Kingdom 33
C Treaties.. 33
D Books and Chapters in Books 33
E Journal Articles 34
F Parliamentary and Government Materials 35
G Reports 36
H Dissertations 36
I Internet Resources 36
J Other Resources 37

II Introduction[1]

On March 15 2019, a lone gunman carried out a terrorist attack in Christchurch, New Zealand. Fifty people were killed and fifty were injured.[2] Attention turned to which of the country's laws could have allowed such an attack. Parliament swiftly tightened gun control,[3] but it was clear the problem was multi-faceted. Ten days after the attack, the Government directed the Royal Commission of Inquiry to investigate and report on what had occurred.[4] The Royal Commission was tasked with investigating three broad areas: "the actions of the individual, the actions of relevant public sector agencies", and most relevant to this essay, "any changes that could prevent such terrorist attacks".[5] Although the Commission's focus was originally intended to be on New Zealand's counter-terrorism laws, a broad issue of "social cohesion, inclusivity and diversity" became apparent as the investigation unfolded.[6] This led to a review of New Zealand's existing hate speech laws.

In December 2020, The Royal Commission released a report detailing their findings which contained 44 recommendations,[7] all of which the Government agreed to "in principle".[8] Four of these recommendations concerned "hate speech" and "hate crime".[9] In June 2021, the Minister of Justice, the Hon Kris Faafoi MP, released a summary of

1 With thanks to my supervisor, Dr Eddie Clark, whose support and guidance has been invaluable to this paper.

2 "Christchurch mosque terror attacks day 8: What you need to know" (22 March 2019) Radio New Zealand <rnz.co.nz>.

3 "The Christchurch mosque attacks: how Parliament responded" (11 March 2021) New Zealand Parliament <www.parliament.nz>.

4 Royal Commission of Inquiry *Ko tō tātou kāinga tēnei: Royal Commission of Inquiry into the terrorist attack on Christchurch masjidain on 15 March 2019* (December 2020) vol 1, Executive Summary at [1].

5 At [3].

6 At [43].

7 Royal Commission of Inquiry, above n 4.

8 (8 December 2020) 749 NZPD (Royal Commission of Inquiry into the Terrorist Attack on Christchurch Masjidain on 15 March 2019 – Report, Rt Hon Jacinda Ardern (Prime Minister)).

9 Note the distinction between hate speech and hate crimes, the latter being irrelevant to the proposal this essay addresses. Whilst hate speech is an independent offence, hate crimes are not. The latter applies when hatred is the motivation for an offence, for example, the targeted assault of a minority *because* of their race. The hateful motivation aggravates an offence to which, accordingly, a more serious penalty will attach. This is reflected in New Zealand legislation by s 9(h) of the Sentencing Act 2002. See also John Ip "Debating New Zealand's Hate Crime Legislation: Theory and Practice" (2005) 21 NZULR 575 at 575.

proposals to implement these recommendations.[10] This essay will concern the proposal to replace s 131 of the Human Rights Act 1993 (HRA), the existing criminal provision which penalises hate speech, with a new provision in the Crimes Act 1961.[11] I will analyse the merit of this proposal, including the harm it seeks to address and whether the proposed provision does so in a justified manner.

Part III summarises the proposal, examining why s 131 was deemed inadequate and outlining the key changes in the new provision: the broadening of the means of communication; the changes in terminology; and the removal of the "likely to" result element. Part IV analyses hate speech, the conduct the proposal seeks to better address, and whether the harm it causes justifies an infringement upon freedom of expression. It will argue a reform of s 131 is justified by this harm, however, the resulting legislation must be narrowly worded to ensure a minimal infringement upon the freedom. This provides normative guidance for part V, which analyses the changes in the proposal. It will argue that while some changes are warranted, the provision requires more definitional clarity, particularly regarding "hatred". Further, the removal of the "likely to" result element should be reversed. I will conclude that the proposal has merit but requires more work to guarantee a narrow application.

III The Proposal

In this part, I will outline the provision that the Minister of Justice proposed to insert into the Crimes Act as a replacement for s 131 of the HRA. I will first summarise the rationale for replacing s 131, as this is important to understanding the changes in the proposal.

A Rationale for Replacing s 131

Section 131 of the HRA, the current New Zealand criminal hate speech provision, states that:

> **131 Public Incitement of Hatred**
>
> 1) Every person commits an offence and is liable on conviction to imprisonment for a term not exceeding 3 months or to a fine not exceeding $7,000 who, with intent to excite hostility or ill-will against, or bring into contempt or ridicule, any group of

[10] Ministry of Justice *Proposals Against Incitement of Hatred and Discrimination* (June 2021).

[11] At 18–19.

> persons in New Zealand on the ground of the colour, race, or ethnic or national origins of that group of persons,—
>
> (a) publishes or distributes written matter which is threatening, abusive, or insulting, or broadcasts by means of radio or television words which are threatening, abusive, or insulting; or
>
> (b) uses in any public place (as defined in section 2(1) of the Summary Offences Act 1981), or within the hearing of persons in any such public place, or at any meeting to which the public are invited or have access, words which are threatening, abusive, or insulting,—
>
> being matter or words likely to excite hostility or ill-will against, or bring into contempt or ridicule, any such group of persons in New Zealand on the ground of the colour, race, or ethnic or national origins of that group of persons.

The core rationale for replacing s 131 was the supposed inadequacy of the provision. The Royal Commission found this to be the case for several reasons. First, it is impracticable to enforce — there have only been two prosecutions under the provision.[12] Section 61 of the HRA, the almost identical civil equivalent of s 131,[13] bears a similar story: there have been only two claims under it.[14] Although this is due in part to the fact that prosecution under s 131 requires the Attorney-General's consent,[15] the Commission viewed the wording of s 131 as overly broad, setting an ostensibly low liability threshold.[16] Because s 131 impinges on freedom of expression — a vitally important right guaranteed by the New Zealand Bill of Rights Act 1990 (NZBORA) — the judiciary is tasked with formulating their own interpretation of the provision to better protect the right.[17] As such, a significant degree of uncertainty had been created. This, the Commission believed, was illustrated by New Zealand's most significant hate speech case: *Wall v Fairfax*.[18]

---

12 *King-Ansell v Police* [1979] 2 NZLR 513 (CA), note that this was a prosecution under s 25 of the Race Relations Act 1971, which was replaced by s 131 of the HRA, however s 131 used the same terms as its predecessor. Another individual has recently pleaded guilty to a charge under s 131 of the HRA, however the sentencing notes are not yet available (to the best of my knowledge): see Benn Bathgate "Man who went 'down the rabbit hole' sentenced for video that incited hate against Māori" (22 April 2022) Stuff <stuff.co.nz>.

13 The key difference being that s 61 does not require intention, whereas s 131 does.

14 *Wall v Fairfax* [2018] NZHC 104, [2018] 2 NZLR 471; and *Proceedings Commissioner v Archer* (1996) 3 HRNZ 123 (CRT). See Royal Commission of Inquiry, above n 4, vol 4, pt 9, ch 4 "Hate crime and hate speech" at [34].

15 Human Rights Act, s 132.

16 Royal Commission of Inquiry, above n 14, at [39].

17 At [39].

18 *Wall v Fairfax*, above n 14.

Although the case was brought under s 61 of the HRA, the almost identical wording of s 131 means it is equally applicable.[19] The defendant in *Wall* had published cartoons depicting Māori and Pasifika parents as lazy, neglectful, gluttonous smokers and drinkers.[20] The issue was whether s 61 was engaged, meaning the cartoons must have been likely to bring Māori and Pasifika into contempt or excite hostility against them.[21] The Court held that s 61 only applied to "relatively egregious examples of expression".[22] This standard, the Royal Commission stated, was an unsatisfactory test for the imposition of criminal liability.[23] The wording of s 131 was also deemed flawed because it does not apply to all forms of communication, instances of online hate speech, for example.[24]

1 Amending s 131

It could be questioned why s 131 would need to be replaced with a provision in the Crimes Act, instead of merely amending it. In answer to this, the Minister of Justice stated its insertion into the Crimes Act would signal that it is a serious offence.[25] This reflects the reasoning of the Royal Commission, who believed the offence's current location in the HRA means it functions more as a value statement.[26] The Crimes Act, however, lists the most serious offences in New Zealand, and the provision's relocation to it would signal this.[27] To further this signalling effect, the Minister of Justice sought to increase the penalty of s 131 from a maximum of three months' imprisonment or a fine not exceeding $7000, to a maximum of three years' imprisonment or a fine not exceeding $50,000.[28]

B The Proposed Changes

The proposed Crimes Act provision addresses the supposed flaws of s 131 of the HRA. Under the new provision, it would be a crime to:[29]

1. intentionally incite/stir up, maintain or normalise hatred

---

19 Royal Commission of Inquiry, above n 14, at [38].
20 *Wall v Fairfax*, above n 14, at [13].
21 At [44].
22 At [56].
23 Royal Commission of Inquiry, above n 14, at [38].
24 At [52].
25 Ministry of Justice, above n 10, at 19.
26 Royal Commission of Inquiry, above n 14, at [34].
27 At [55].
28 At 32.
29 Ministry of Justice, above n 10, at 30.

2. against any group protected from discrimination by section 21 of the Human Rights Act
3. through threatening, abusive or insulting communications, including inciting violence
4. made by any means.

There are three areas of change I will address: the broadening of the means of communication; the change in terminology; and the removal of the "likely to" result element.

1 Means of Communication

As outlined above, s 131 of the HRA only applies if the communication is written; are words which are broadcasted by radio or television; or spoken within a public place.[30] To address this, the proposed provision would apply to "communications … made by any means".[31] Therefore, the provision would capture hate speech communications made online. This was a key goal of the Royal Commission, who were concerned by the prevalence of online hate speech.[32]

2 Terminology

Although the requirement of intention remains, the proposed provision contains numerous changes to the terminology of s 131.

(a) "Hatred"

Instead of "excit[ing] hostility or ill-will against, or bring[ing] into contempt or ridicule [the protected class]", a defendant would now have to "incite/stir up, maintain or normalise hatred".[33] Thus, "hostility", "contempt" and "ridicule" are replaced by one catch-all term, "hatred". The Royal Commission believed this change would create the judicial certainty currently lacking in s 131, as "hatred" is what they understood to be a precise term "impl[ying] extreme dislike or disgust, including an emotional aversion".[34] If "hatred" was used, the Commission argued, it would narrow the scope of the offence

---

30 Section 131(1)(a)–(b).

31 Ministry of Justice, above n 10, at 30.

32 Royal Commission of Inquiry, above n 14, at [13], [52] and [81]; and Royal Commission of Inquiry *Hate speech and hate crime related legislation* (November 2020) [*Companion Paper*] at [15]–[19] and [40].

33 Ministry of Justice, above n 10, at 30.

34 Royal Commission of Inquiry, above n 14, at [43].

and therefore prevent the judiciary from applying imprecise interpretations, such as in *Wall v Fairfax*.[35]

(b) "Stir up", "Maintain" and "Normalise"

Under s 131, the defendant must "excite" hostility against the protected class. The proposed provision changes this by replacing "excite" with "incite/stir up, maintain or normalise [hatred]".[36] This addresses several of the Commission's concerns, the first of which is the use of "excite". The Commission believed this suggested causation, in that "excite" could only be engaged if the "hostility or ill-will" did not previously exist, or if such hostility was enhanced or increased.[37] This was an issue if the defendant was "preaching to the converted",[38] for example, if the leader of a Neo-Nazi organisation was espousing racist vitriol which did not increase the group's hatred against the race because that hatred already reflected the group's feelings. In addition, the proposal seeks to replace "excite" with "incite/stir up". The Minister of Justice stated it was still undecided which out of "incite" or "stir up" would be used.[39] The Royal Commission recommended the latter, as it reflects corresponding United Kingdom legislation.[40] This is indeed the case — s 18(1) of the Public Order Act 1986 requires the defendant intended to "stir up racial hatred".

(c) Intention

The mens rea standard of intention remains unchanged. In other words, a defendant under the proposed provision must intentionally incite hatred against the target group. The Minister of Justice believed this was justified by the increased penalty carried by the proposed provision.[41] The Royal Commission stated that retaining intention would set a high liability threshold, helping to alleviate concerns about freedom of expression.[42]

35 Royal Commission of Inquiry, above n 14, at [43].

36 Ministry of Justice, above n 10, at 30.

37 At [40].

38 At [40]; and Ministry of Justice, above n 10, at 19.

39 At 19.

40 At [41].

41 At 30.

42 At [51].

3 *Removal of the "likely to" Result Element*

Significantly, the proposed provision removes the requirement that the communication must be "likely to" incite, maintain or normalise hatred.[43] This is what the Court in *Wall v Fairfax* deemed to be an "objective effects-based test",[44] and assessed "… whether a reasonable person, aware of the context and circumstances surrounding the expression, would view it as likely to expose the protected group to the identified consequences".[45] Under the proposed provision, however, the offence would be committed if the defendant:[46]

a) Intentionally incited hatred;
b) Against a protected group; and
c) Did so through threatening, abusive or insulting communications (made by any means).

Thus, the focus shifts from whether hatred is likely to result from the communication to the act of inciting hatred itself. The Minister of Justice adopted the Royal Commission's approach, who believed the "likely to" element was unnecessary.[47] This was because if it could be proven that the defendant was responsible for a threatening communication through which they intended to incite hatred, whether such hatred was likely to result has "little or no bearing on whether the conduct is sufficiently culpable to justify in a charge".[48]

C *Summary*

The proposed provision is intended to be a functional replacement of s 131, better addressing the harm of hate speech whilst maintaining a high liability threshold. It also sought to reflect the seriousness of hate speech, demonstrated by its shift into the Crimes Act and an increased penalty. The key reasons for this reform were the perceived ineffectiveness of s 131 (demonstrated by the limited cases in which it has been engaged)

---

43 Ministry of Justice, above n 10, at 30–31.

44 *Wall v Fairfax*, above n 14, At [47].

45 At [51].

46 Ministry of Justice, above n 10, at 30.

47 Ministry of Justice, above n 10, at 31.

48 Royal Commission of Inquiry, above n 14, at [46].

and an increased recognition of the harm flowing from hate speech. The next part discusses this latter reason.

IV Hate Speech

The Royal Commission believed hate speech was increasingly socially unacceptable and a contributing factor to the issue of social cohesion,[49] and intended the provision to better address the harm flowing from it.[50] In this part, therefore, I will first analyse what "hate speech" means, and then turn to its harm. I will discuss whether this harm justifies a legal response, despite its infringement upon freedom of expression. This provides some normative standards for hate speech legislation,[51] which I will analyse the proposal against in part V.

A Defining Hate Speech

Hate speech is a difficult term to define,[52] illustrated by the Minister of Justice's definition: "[hate speech is] speech that attacks an individual or group based on a common characteristic, for example ethnicity, religion, or sexuality".[53] Although this is an acceptable starting point, a true normative definition of hate speech is more nuanced. First, it is not confined to speech. As Bromell explains, any form of communication can incite hostility, whether "spoken (speech), written, mimed, memed, graffitied, cartooned or tweeted".[54] In fact, non-spoken forms are arguably more troubling: a published word or image lingers more than speech.[55] Although this has led many to caution against using

---

49 Royal Commission of Inquiry, above n 14, at [4].

50 At [13].

51 It is uncontroversial to class the provision as hate speech legislation. Hate speech is both the conduct which the provision seeks to address and falls within Alexander Brown's fourth cluster of hate speech laws, that being laws which prohibit the intentional incitement of hatred, via "speech or other expressive conduct", toward a protected group: Alexander Brown *Hate Speech Law: A Philosophical Examination* (Routledge, New York and Oxford, 2015) at 26.

52 See Bhikhu Parekh "Is There a Case for Banning Hate Speech?" in Michael Herz and Peter Molnar (eds) *The Content and Context of Hate Speech: Rethinking Regulation and Responses* (Cambridge University Press, New York, 2012) 37 at 40; and Elizabeth Macpherson "Regulating Hate Speech in New Zealand" (LLB (Hons) Research Paper, Victoria University of Wellington, 2003) at 3.

53 Ministry of Justice, above n 10, at 9.

54 David Bromell *'Hate Speech': Defining the Problem and Some Key Terms* (Institute for Governance and Policy Studies, Victoria University of Wellington, Working Paper 21/03, March 2021) at 13–14.

55 Bromell, above n 54, at 14 citing Jeremy Waldron *The Harm in Hate Speech* (Harvard University Press, 2012) at 37–38.

the term hate *speech*,[56] the Royal Commission agreed they could "see no good reason why there should be restrictions based on how hate speech is communicated".[57] Second, although hate speech may convey hateful speech, it is distinct from the mere expression of hatred. It is a narrow class of communication and needs to be distinguished from insulting or offensive speech when legislating against it.[58] Seeking to formulate a definition, Bhikhu Parekh lists three core elements of hate speech. His definition is useful as a normative standard because it reflects elements of s 131 and the proposal.[59]

The first element of hate speech is that:[60]

> … it is directed against a specified or easily identifiable individual or, more commonly, a group of individuals based on an arbitrary and normatively irrelevant feature.

This is reflected in the proposal, as the communication must be targeted against "any group protected from discrimination by s 21 of the Human Rights Act".[61] This is a fundamental element, as it helps distinguish hate speech from mere insulting or offensive communications. As Herz and Molnar write, "[t]elling an ex-lover 'I hate you' might be an expression of hate, but it is not hate speech".[62] The communication must be targeted at a group, or an individual *because* they belong to that group.[63] Thus, hate speech is still distinct from insulting an individual who is coincidentally part of that group.

Parekh's second element is that:[64]

> … [h]ate speech stigmati[s]es the target group by *implicitly or explicitly* ascribing to it qualities widely regarded as highly undesirable.

---

56 See Bromell, above n 54, at 14.

57 Royal Commission of Inquiry, above n 14, at [52].

58 Parekh, above n 47, at 40.

59 More broadly, his definition reflects the "incitement to hatred" cluster of hate speech laws which Brown identifies. The two provisions both fall within this cluster. See Brown, above n 51, at 26.

60 At 40.

61 Ministry of Justice, above n 10, at 30.

62 Michael Herz and Peter Molnar "Introduction" in *The Content and Context of Hate Speech: Rethinking Regulation and Responses* (Cambridge University Press, New York, 2012) 1 at 3.

63 See Waldron, above n 55, at 37; and Bromell, above n 54, at 12.

64 At 41 (emphasis added).

This expands on what it may mean to incite "hatred" under the proposed provision and reflects the s 131 requirement that the defendant intended to "excite hostility ill-will against, or bring into contempt or ridicule [the target group]".[65] As this element recognises, hate speech may be blatant. For example, the defendant in the Canadian case of *R v Topham* was convicted for describing Jewish people as having an inherent "war-soul" which presented a danger to "human safety and security".[66] Hate speech can also be implicit. Another anti-Semite was convicted in Canada for statements such as:[67]

> I have a simple solution to guarantee eternal world peace: give every country, no matter how small, ONE NUKE! Have the nuke robotically controlled to auto-launch if the robot determines the country is being invaded. Then point every one of them at Tel Aviv, Israel.

In this example, the defendant implies that Jewish people are, and will continue to be, responsible for all warfare. If any country is invaded, therefore, the capital of Israel should be destroyed in retaliation. This sanctioning of the mass murder of Jewish people aligns with Parekh's third element of hate speech, which reflects the importance of its effect:[68]

> … because of its negative qualities, the target group is viewed as an undesirable presence and a legitimate object of hostility. It cannot be trusted to be a loyal member of society and presents a threat to its stability and well-being. [Thus, society] … may legitimately exterminate or expel the target group, [or] discriminate against and tolerate it as an unavoidable evil … Thus, hate speech encourages and purports to justify discrimination.

On the more extreme end of the continuum, hate speech may act as a call for immediate violence against the target group. However, this is not an absolute requirement: a normative definition of hate speech,[69] the proposal and s 131 do not require immediacy.[70]

---

65 Human Rights Act, s 131(1).

66 *R v Topham* [2017] BCSC 259 at [55].

67 *R v Sears* [2019] ONCJ 104 at 23.

68 At 41.

69 Parekh, above n 52, at 41.

70 But see Peter Molnar "Responding to 'Hate Speech' with Art, Education, and the Imminent Danger Test" in Michael Herz and Peter Molnar (eds) *The Content and Context of Hate Speech: Rethinking Regulation and Responses* (Cambridge University Press, New York, 2012) 183, who argues that hate speech legislation should require a broadly defined "imminent danger test". This would mean the communication must cause imminent danger to be prohibited, but does not require that the communicator intended to incite danger: at 195–196.

On the less extreme, but far more common end, hate speech creates a culture in which discrimination against the target group is justified.[71] This is manifest in s 131(1) of the HRA, as the communication must be "likely to" bring into effect the defendant's intent. As I will discuss in part V, the proposal omits this requirement. The effect of hate speech is important, as it further individualises it as a unique class of communications and provides the basis for legislating against it, which I will expand on below.

B The Harm of Hate Speech

The overarching motivation for the provision's introduction was to better address the harm of hate speech. The Royal Commission provided examples of research that correlated an increased presence of hate speech with physical hate crimes,[72] and a general degradation of social cohesion.[73] In this sub-part, I will analyse the harm of hate speech, and whether that harm justifies legal prohibition.

From an abstract standpoint, hate speech creates an environment where discrimination and violence against the target group are legitimised. Waldron argues that, in a well-ordered society, each individual deserves a sense of security. Hate speech, a "slow-acting poison", undermines this, "accumulating here and there", until the communicator's view of the target group is legitimised and commonly held.[74] Dickson CJ of the Canadian Supreme Court made a similar point in *R v Keegstra*:[75]

> A … harmful effect of hate propaganda … is its influence upon society at large. It is … not inconceivable that the active dissemination of hate propaganda can attract individuals to its cause, and in the process create serious discord between various cultural groups in society. … Even if the message of hate propaganda is outwardly rejected, there is evidence that its premise of racial or religious inferiority may persist in a recipient's mind as an idea that holds some truth …

Moreover, hate speech harms members of the target group's quality of life. Dickson CJ also described how:[76]

---

71 Parekh, above n 52, at 41.

72 Royal Commission of Inquiry, above n 14, at [13].

73 At [9].

74 Waldron, above n 55, at 4.

75 *R v Keegstra* [1990] 3 SCR 697 at 747–748.

76 At 746–747.

> The derision, hostility and abuse encouraged by hate propaganda … have a severely negative impact on the individual's sense of self-worth and acceptance. This impact may cause target group members to take drastic measures in reaction, perhaps avoiding activities which bring them into contact with non-group members or adopting attitudes and postures directed towards blending in with the majority.

The harm of hate speech can therefore be viewed as two-fold: the diminishing of an individual's self-worth who belongs to the target group; and a culture where discrimination and violence against the group's members are eventually legitimised and commonplace. Thus, as Waldron argues, a well-ordered society would legislate against hate speech. Not only would it limit the actual dissemination of hate speech and its subsequent harm, but it would give assurances to those otherwise targeted of their dignity and security in society.[77] This elaborates on the Royal Commission's motivation of strengthening social cohesion through the provision.[78] Although such a motivation may appear overly broad and idealistic, the effect of hate speech is indeed to degrade such cohesion.

Admittedly, the harm outlined above may seem too abstract. It is harder to quantify the extent to which hate speech creates the culture which Waldron and Dickson CJ describe, than it is, for example, to measure the number of physical assaults per year. This has led some to characterise the effect of hate speech as nothing more than personal, emotional offence. Stephen Guest, an opponent of the proposal, argues that the only harm which arises from hate speech is "hurt feelings, feelings of being wronged, being dismayed by what other people say or do, or even being shocked, and suffering as a result".[79] He proposes hate speech legislation is unjustified, as in most cases, these effects are merely "part of the human condition", and there is no right to not be emotionally harmed, nor to prevent a particular view from being expressed to us.[80]

I disagree with this sentiment. Aside from a minimisation of the personal harm an individual who is a member of the target group may feel, there is strong evidence to support the conclusion that hate speech does indeed create the culture I described above.

---

77 Waldron, above n 55, at 82–85.

78 Royal Commission of Inquiry, above n 14, at [1].

79 Stephen Guest "Why Hate has no Place in the Criminal Law" [2021] NZLJ 371 at 372.

80 At 372.

The Royal Commission referred to a 2020 study conducted in the United Kingdom to support this point.[81] The authors of this study sought to determine the link between instances of online hate speech and hate crimes. To do this, they compared instances of hate crime in London with Twitter records over an eight-month period.[82] Their findings indicated a "consistent positive association between Twitter hate speech targeting race and religion and offline racially and religiously aggravated offences in London".[83] Unfortunately, the nexus between hate speech and real-world harm is equally applicable to New Zealand. Moments before carrying out the Christchurch attack, the terrorist posted an online message, stating his attack was a "real life effort post" — in other words, an attempt to migrate the online culture of anti-Muslim rhetoric into reality.[84] Hate speech, therefore, increases and legitimises violence against a protected group. It is hard to imagine a culture sanctioning the group's discrimination would not be a corollary of this.

Hate speech poses a very real threat to New Zealand citizens. Following the attack, the Federation of Islamic Associations of New Zealand (FIANZ) carried out a series of hui to determine areas of concern for Muslims. New Zealand Muslims considered hate speech to be the fourth-most important area of concern, behind Islamophobia, social cohesion and gun licencing.[85] In a similar series of hui carried out in 2020 by the Department of the Prime Minister and Cabinet, "many spoke of the importance of the reform of hate speech legislation, as they felt that it is critical for increasing a sense of safety".[86] Similarly, 52 per cent of Muslim respondents and 15 per cent of total respondents in a 2019 New Zealand survey stated they had experienced hate speech through digital communications in the last year.[87] The prevalence of hate speech suffered by target

[81] Royal Commission of Inquiry, above n 14, at [13].

[82] Matthew Williams, Pete Burnap, Amir Javed, Han Liu and Sefa Ozalp "Hate in the Machine: Anti-Black and Anti-Muslim Social Media Posts as Predictors of Offline Racially and Religiously Aggravated Crime" (2020) 60(1) Br J Criminol 93 at 94.

[83] At 111.

[84] At 97.

[85] "Preliminary Submission on the Proposed Hate Speech Legislation" The Federation of Islamic Associations of New Zealand (Inc.) <www.fianz.com> at 4.

[86] At 4, citing Department of the Prime Minister and Cabinet *Report on Community Hui Held in Response to the Royal Commission into the Terrorist Attack on Christchurch Mosques on 15 March 2019* (March 2021) at 5.

[87] Edgar Pacheco and Neil Melhuish "Measuring Trends in Online Hate Speech Victimisation and Exposure, and Attitudes in New Zealand" (December 2019) Netsafe <www.netsafe.org.nz> at 5–6. This

groups indicates the culture exists in New Zealand, and distinguishes it from isolated incidents where the key effect may be individual, emotional harm. Waldron argues that when this harm begins to become a reality — as I argue it has in New Zealand — legal regulation of hate speech is justified.[88] For New Zealand, the form of this regulation must be reformatory, as the growing prevalence of hate speech and its subsequent harm indicates s 131 is unfit to address it.

C Freedom of Expression

While the harm caused by hate speech may be real, there is strong resistance to its legal regulation. The main opposition is founded upon such legislation's infringement upon freedom of expression.[89] Hate speech laws necessarily impinge on the freedom because it restricts what a citizen can and cannot communicate. This is important in New Zealand, where freedom of expression is protected by NZBORA. Section 14 states: "Everyone has the right to freedom of expression, including the freedom to … impart information and opinions *of any kind* in any form" (emphasis added). Thus, ostensibly, hate speech is protected by s 14. As NZBORA recognises, however, rights must be balanced against each other. Section 5 allows for the rights and freedoms contained in the Act to be subject to reasonable limitations if they are "demonstrably justified in a free and democratic society". Therefore, freedom of expression is not an absolute bar on hate speech legislation, and the arguments in opposition to such laws must be weighed against the harm of hate speech.[90]

Whether freedom of expression should preclude hate speech laws is a complex issue, as it is considered by many to be an inviolable, "almost human right and the cornerstone of

---

survey is useful as the definition of hate speech provided to respondends is similar to the one I have outlined, see at 4.

[88] Waldron, above n 55, at 68.

[89] See for example Guest, above n 79; Katharine Gelber *Speaking Back: The Free Speech Versus Hate Speech Debate* (John Benjamins Publishing Company, Amsterdam and Philadelphia, 2002), who argues that to balance freedom of expression and hate speech, a policy of "speaking back" should be implemented, enabling victims of hate speech to use their freedom of expression to counter hate speech communications; and Michael Conklin "The Overlooked Benefits of 'Hate Speech': Not Just the Lesser of Two Evils" (2020) 60 S Tex Law Rev 687.

[90] A point observed by the Royal Commission of Inquiry: above n 14, at [8].

democracy".[91] Under this view, freedom of expression is a supreme right which can only be limited by law when there is a threat of immediate violence.[92] Anne Flahvin identifies two core justifications for granting freedom of expression this inviolable status: its relation to democratic self-governance; and the marketplace of ideas rationale.[93] I will examine these justifications below.

1 Democratic Self-Governance

First, freedom of expression is seen as essential to democratic self-governance. Self-governance is the concept that the authority of government institutions in a democratic society arises from the individuals who make up that society,[94] reflected in the citizen's right to vote.[95] Essential to self-governance is the citizen's exposure to a diverse range of views on a particular issue, without the government censoring certain ones. Freedom of expression guarantees this and has allowed social movements for civil and women's rights, for example, to enhance equality.[96] An off-shoot of self-governance is the importance of individual autonomy.[97] This is a fundamental principle of democratic theory: Edwin Baker argues that a state's legitimacy reflects the extent its laws regard its citizens as autonomous.[98] As such, the law should never deny an individual their autonomy — including their freedom of expression. From this lens, Baker believes legal prohibitions on hate speech should be "generally impermissible".[99] Although under the principle of autonomy a guarantee of one individual's autonomy should never impinge on the autonomy of another,[100]

---

[91] Juliet Moses "Hate Speech: Competing Rights to Freedom of Expression" (1996) 8 AULR 185 at 189–190. For a modern example see Joe Dryden "Protecting Diverse Thought in the Free Marketplace of Ideas: Conservatism and Free Speech in Higher Education" (2018) 23(1) Tex Rev L & Pol 229 at 263.

[92] On the immediacy point see Peter Molnar, above n 70.

[93] Anne Flahvin "Can Legislation Prohibiting Hate Speech be Justified in Light of Free Speech Principles?" (1995) 18 UNSWLJ 327 at 330.

[94] Colin Bird "The Possibility of Self-Government" (2000) 94(3) APSR 563 at 563–564.

[95] Douglas Fraleigh and Joseph Tuman *Freedom of Expression in the Marketplace of Ideas* (SAGE, London, 2011) at 8.

[96] Fraleigh and Tuman, above n 95, at 8.

[97] Bird, above n 94, at 563.

[98] Edwin Baker "Hate Speech" in Michael Herz and Peter Molnar (eds) *The Content and Context of Hate Speech: Rethinking Regulation and Responses* (Cambridge University Press, New York, 2012) 57 at 64.

[99] At 64.

[100] At 64.

> … [r]estrictions on … hate speech violate the speaker's formal autonomy, whereas [the communicated hate speech] does not interfere with or contradict anyone else's formal autonomy even if such speech does cause injuries that sometimes include undermining others' substantive autonomy.

In other words, hate speech regulations restrict the speaker's formal autonomy — their capacity for individual choice, which is here used to freely express their views — whereas communicated hate speech does not directly restrict the recipient's. Although autonomy is important, Baker's argument is inconclusive. It still leaves room to debate why the speaker's formal autonomy trumps the target's substantive authority, their actual capacity to lead a "meaningful, self-directed life".[101] When injuries undermining substantive authority have become commonplace, I argue a narrow limitation on formal authority is justified.

2 *Marketplace of Ideas*

Secondly, freedom of expression allows for a "marketplace of ideas". This concept first assumes the necessity and good of truth.[102] Freedom of expression is essential to the truth because it creates an environment where views and opinions can be freely traded without fear of legal repercussions. In such a system, the belief is if all views are expressed, they will compete against each other until the truth surfaces. This metaphor was coined by Holmes J in 1919, who wrote in the United States Supreme Court that:[103]

> … the ultimate good desired is better reached by free trade in ideas — that the best test of truth is the power of the thought to get itself accepted in the competition of the market…

The marketplace of ideas is motivated by hindsight, namely the historical impact of censorship. Fraleigh and Tuman point to the fact that ideas now universally accepted as truth were once suppressed.[104] For example, Galileo was forced to renounce his view that the sun was at the centre of the universe because it contradicted the religious authority's beliefs. Although modern Western society is generally unaccepting of government censorship, the marketplace of ideas remains important. The knowledge which grows

[101] At 64.
[102] Fraleigh and Tuman, above n 95, at 10.
[103] Fraleigh and Tuman, above n 95, at 10, citing *Abrams v United States* 250 US 616 (1919) at 630.
[104] At 10–11.

from social interactions is enhanced by "listening to diverse viewpoints with an open mind and contributing our own perspectives to the debate".[105] Opponents of hate speech legislation on this ground argue prohibitions create a "chilling effect" on the marketplace,[106] as silencing offensive views hamstrings the ability for such views to be defeated by the truth.[107] Thus, some argue the correct approach to countering hate speech is with more speech.[108] Katharine Gelber, for example, recognises the harm in hate speech as more than emotional offence but believes the proper legislative response should be policies that empower targets of hate speech to respond.[109] This is an attractive argument, as it justifies a legal response to the harm of hate speech while avoiding an encroachment upon free speech.

3 Harm of Hate Speech as an Infringement on Free Speech

There is a strong argument, however, that the debate does not need to be distilled into hate speech versus free speech. In fact, the effect of hate speech can be understood as undermining freedom of expression. Caroline West argues freedom of expression, in its normative form, requires that opinions are considered — not only voiced. This is especially so in the context of the marketplace of ideas:[110]

> For speech to play a role in facilitating knowledge, audiences must reasonably often be in a position not simply to hear what speakers say, but also to attend to it, to give it some consideration, and to update their beliefs and desires in light of the perceived merits of the information received.

Therefore, a market failure would occur if a speaker was "discouraged from speaking by an unsympathetic speech environment in which what she has to say is likely to be ridiculed".[111] The effect of hate speech is to create this environment, one where any response from the target group is automatically chastised. Imagine an African American

---

105 At 11.

106 Dryden, above n 91, at 243.

107 At 252.

108 See, for example, at 252.

109 Gelber, above n 89, at 9.

110 Caroline West "Words that Silence? Freedom of Expression and Racist Hate Speech" in Ishani Maitra and Mary Kate McGowan (eds) *Speech and Harm: Controversies over Free Speech* (Oxford University Press, Oxford, 2012) 222 at 230.

111 At 228.

giving a speech at a Ku Klux Klan rally. Would we expect the audience to rationally understand their position? Whilst this may be theoretically ideal, it is not realistic. It is unlikely the audience will consider their response, and even more probable that members of the target group will remain silent out of fear. West writes that if they were to respond:[112]

> … they know that … there is a very real chance that they will end up beaten, incarcerated, or dead. In a situation where speaking out would be very costly, only the exceptionally courageous (or foolhardy) are likely to speak. Most people, quite reasonably, will remain silent.

Therefore, whilst a policy empowering targets of hate speech to speak back may make normative sense, it is unrealistic. This shows the debate between hate speech and freedom of expression is more nuanced than its name suggests and may even be understood as balancing the communicator's freedom of expression with that of the recipient.

D The Need for Caution

Freedom of expression is justifiably important. However, it should not be an absolute bar on hate speech legislation, especially when the harm undermines the target's own freedom of expression. The balance between freedom of expression and addressing the harm in hate speech is an important factor in drafting hate speech law, which the Royal Commission observed.[113] I argue that, if a legal prohibition of hate speech is accepted, the guiding principle should be narrowness. Great care needs to be used when drafting such legislation, and definitional clarity should result. This will ensure the infringement upon freedom of expression is minimal, which is fundamental to ensuring the provision is a "demonstrably justified" limitation under s 5 of NZBORA.[114] Per *R v Hansen*, a key element of this inquiry is whether the "limiting measure impair[s] the right or freedom no more than is reasonably necessary for sufficient achievement of its purpose".[115] I argue the harm of hate speech means its legal prohibition is a justified end. However, the proportionality requirement means a narrow offence created by clear terminology is essential.

---

[112] At 234.

[113] Royal Commission of Inquiry, above n 14, at [8].

[114] *R v Hansen* [2007] NZSC 7, [2007] 3 NZLR 1 at [123] per Tipping J.

[115] At [104] per Tipping J.

The unique status of prohibited speech makes this a difficult, but increasingly necessary requirement. The word "crime" typically calls to mind physical acts. Although how those acts may be carried out will inevitably vary, it is generally easier to identify a prohibited act than it is to identify prohibited speech. For example, assault — the application of force to another —[116] is more recognisable than communication which (using the wording of the proposal) incites hatred against a protected group. The task of defining the point at which communication crosses from an insult or distressing remark into the threshold of hate speech is one of the most difficult, but most important, for lawmakers.[117] This is a point the Royal Commission observed:[118]

> The difference between legally criminalised hate speech and the vigorous exercise of the right to express opinions is not easy to capture in legislative language. As well, the more far reaching a law creating hate speech offences, the greater the potential for inconsistency with the right to freedom of expression under section 14 of the New Zealand Bill of Rights Act.

Taking guidance from Bromell, it is important for legislation to distinguish between hate speech and communication that is reprehensible, but undeserving of criminality.[119] For example, communications that:

- Express dislike of a group but do not incite discrimination, hostility or violence against them;
- Demean the beliefs of a religious group, but do not incite hatred against participants of that group;
- Insult an individual who belongs to a protected group, but is limited to insulting that individual, not the group they belong to.

Thus, hate speech legislation must be clear in its wording. When the narrow definition of hate speech is coupled with the inherent ambiguity of language, a broadly worded provision would be unfeasible, setting the bar of criminality too low and encroaching upon free speech more than is reasonably necessary.

---

116 Crimes Act, s 2(1).

117 Herz and Molnar, above n 62, at 3; and Parekh, above n 52, at 53.

118 Royal Commission, above n 14, at [5]; and see Bromell, above n 54, at 11.

119 At 15.

E Summary

I have argued hate speech is a narrow class of communications, as the fact it must be targeted at a group or an individual because they belong to that group, distinguishes it from the mere expression of hatred. The effect of hate speech is also integral to this distinction: the creation of an environment where violence and discrimination against the target group are legitimised. I have argued this environment has become enough of a reality for many New Zealanders to justify reforming s 131, but only if the resulting law reflects the narrowness of hate speech's normative definition. Although hate speech legislation impinges upon freedom of expression, a right considered by many to be inviolable, it can be limited under NZBORA if such legislation is demonstrably justified. This requires a high degree of care and resulting clarity in drafting to ensure the provision has a sufficiently narrow application to satisfy the proportionality requirement.

V Analysis

The harm resulting from hate speech justifies a legal response, and I argue this harm is concerning enough to warrant the provision's shift into the Crimes Act. The nexus between hate speech and real-world harm indicates that s 131 is currently unfit for purpose. I agree with the Royal Commission that this relocation, aided by the increased sentence, will signal the unacceptability of hate speech and the seriousness of its harm.[120] How this provision is realised, however, is important. In this part, I will analyse the proposed provision to determine whether it correctly addresses hate speech. The previous part gives normative guidance to this, especially the guiding principle of narrowness achieved through definitional clarity, to ensure a minimal infringement upon freedom of expression. There are three areas of change I will examine: the means of communication; the replacement of s 131's terminology with "hatred"; and the removal of the "likely to" result element. While I will not examine the insertion of "stir up", "maintain" and "normalise" separately, the latter two terms are important to my discussion of the removal of the result element.[121]

---

[120] Royal Commission of Inquiry, above n 14, at [55].

[121] Regarding the insertion of "stir up", it could be argued that "incite" is the better option as it has been interpreted as a catch-all term, inclusive of the former: *Young v Cassells* (1914) 33 NZLR 852 (HC) at 854; and affirmed more recently in *Sunol v Collier (No 2)* [2012] NSWCA 44, (2012) 289 ALR 128 at [41(a)]. See also Letter from Tiani Epati (President of the New Zealand Law Society) to the Ministry of

A Means of Communication

First, the broadening of the means of communication to include "any means" — most importantly, electronic —[122] is entirely justified. Not only is a normative definition of hate speech unconfined by the means in which it is communicated, the rapid increase of online hate speech and the harm it has been proved to cause indicates the s 131 restriction is a legislative oversight. "Any means" will also capture images and videos, whose lasting presence causes the same, if not more, harm than spoken and written hate speech.[123] If nothing else, s 131 should be amended to effect this change.

B "Hatred"

Next, the replacement of "hostility", "contempt" and "ridicule" with "hatred".[124] There is understandable rationale for this change. The Court in *Wall v Fairfax* considered "hatred" as essentially synonymous with the former terms.[125] It cited the Supreme Court of Canada, which held that, in a human rights context, "hatred" carried the same connotation as "contempt".[126] Moreover, "hatred" is used in similar international legislation prohibiting hate speech.[127] In Canada, it is present in the s 319(1) of the Criminal Code:[128]

> **Public Incitement of Hatred**
>
> 319(1) Everyone who, by communication statements in any public place, *incites hatred ...*

In Germany:[129]

> **Incitement of Masses**
>
> (1) Whoever, in a manner which is suitable for causing a disturbance of the public peace, 1. *incites hatred ...*

Justice regarding the proposals against incitement of hatred and discrimination (5 August 2021) at [4.4] and [4.5]. This is a relatively minor point, however, and underserving of a separate discussion.

122 Ministry of Justice, above n 10, at 30.

123 Bromell, above n 54, at 14 citing Waldron, above n 55, at 37–38.

124 Ministry of Justice, above n 10, at 30.

125 At [52].

126 At [54], citing *Saskatchewan Human Rights Commission v Whatcott* 2013 SCC 11, [2013] 1 SCR 467 at [43].

127 Bromell, above n 54, at 13.

128 Criminal Code RSC 1985 c C-46, s 319(1) (emphasis added).

129 Strafgesetzbuch [German Criminal Code] 1872, s 130(1) (emphasis added).

And, most relevantly, considering the Royal Commission's intent on bringing s 131 more in line with United Kingdom legislation,[130] s 18(1) of the Public Order Act requires the defendant intended to "stir up racial *hatred*".[131] The use of "hatred" also accords with the international obligations under which New Zealand must implement hate speech legislation,[132] particularly art 4(a) of the International Convention on the Elimination of all forms of Racial Discrimination, which requires states to: "declare an offence punishable by law all dissemination of ideas based on racial superiority or *hatred*".[133]

The mere inclusion of "hatred", however, is not enough. It must be properly defined, which the proposal lacks. The definition of "extreme dislike or disgust, including an emotional aversion" the Royal Commission provides is inadequate.[134] "Hatred" in the context of hate speech legislation is unique. Parekh writes that it implies a "wish to harm or destroy the target group, a silent or vocal and a passive or active declaration of war against it".[135] An "emotional aversion" suggests the offence is concerned with the emotion of hatred. A better definition is needed to ensure the focus remains on the harm of hate speech.[136] As I indicated above, a sharp line must be drawn between hate speech and communications that are merely insulting. Otherwise, the provision risks an unintentionally wide application, presenting an unjustified encroachment upon freedom of expression. If the Minister of Justice intended to cement the seriousness of hate speech with the provision's relocation to the Crimes Act and an increased penalty,[137] the same seriousness must be applied to drafting.

Moreover, it is doubtful the main objective of using "hatred" — the creation of judicial certainty — will be achieved. The Court in *Wall v Fairfax*, the case the Royal Commission cited as demonstrative of the certainty issues s 131 created, applied the terminology of s

---

130 Royal Commission of Inquiry, above n 14, at [41].

131 Section 18(1)(a).

132 Royal Commission of Inquiry, above n 14, at [26]–[30]. See also the New Zealand Law Society, who expressed concern on this point: Epati, above n 121, at [6.1.2].

133 International Convention on the Elimination of All Forms of Racial Discrimination 660 UNTS 195 (opened for signature 21 December 1965, entered into force 4 January 1969) (emphasis added).

134 Royal Commission of Inquiry, above n 14, at [42].

135 Parekh, above n 52, at 40.

136 Bromell, above n 54, at 16.

137 Ministry of Justice, above n 10, at 19–20.

61/131 as if it was synonymous with “hatred”.[138] It was upon this basis they created the “relatively egregious” test which the Commission viewed as manifest of the problems with s 131.[139] Thus, the outcome would have been no different if “hatred” was used by s 61/131. This shows that, if the provision was to use “hatred” instead of the current s 131 terminology, a clear definition must be provided. The mere inclusion of the term will not act as a silver bullet to defeat judicial uncertainty and interpretation.

C Removal of the Result Element

Finally, the removal of the requirement that the communication must be “likely to” effect the defendant’s intention.[140] The removal of this objective element is the most flawed aspect of the proposal, as it has a great bearing on the structure of the offence. The lack of detail from both the Minister of Justice and the Royal Commission as to how the offence must be proven without this element is concerning. Whereas under s 131 the act was to publish communications with an intent to excite hostility and that intention was likely to be carried out; under the new provision, the act is to intentionally incite hatred through the communications. There are two possible implications of this.

First, it suggests a requirement that the Crown must prove hatred was *actually* incited against the target group. If this is so, then the threshold of liability would, ostensibly, be greatly increased. Practically, however, it would invoke an unreasonable degree of subjectivity. As Juliet Moses explains:[141]

> Requiring actual proof would make vast amounts of expert necessary and lengthen trials. Further, the test of actual effects would not fully remove the appearance of politisation. Indeed it would probably involve more speculation and subjectivity since it would be necessary to assess the [recipients’] prior feelings towards the target group and any changes following the comments. Thus, to a large extent, a conviction would depend on the nature and background of the recipients and not the words themselves. A more rational, well-educated group may be less influenced.

---

138 At [52]–[55].

139 *Wall v Fairfax*, above n 14, at [56].

140 Ministry of Justice, above n 10, at 30.

141 Moses, above n 91, at 188–189.

Thus, a requirement of actual proof would have unintended consequences, and would, in fact, create the judicial uncertainty the Royal Commission sought to eliminate.

It is unlikely this requirement was the intention, however, as to achieve it the current wording of "likely to" could merely be changed to "did". It is more probable the Minister of Justice expected an intention to incite hatred through the communication would suffice, in that hatred does not actually need to be incited.[142] But this is equally as concerning. From a broad perspective, it unjustifiably removes the requirement of an effect. If the reason for criminalising hate speech is its harm, then, especially considering the tax on freedom of expression, surely the communication must be likely to bring about that effect. Parekh does not believe hate speech should only be defined as communication that has a likely effect.[143] However, Parekh's issue lies in the normative definition of hate speech, not its legal prohibition. He writes that: "what matters is its content — what it says about an individual or a group — *and its long-term effect on the group and the wider society*".[144] So, even if the definition of hate speech may not be limited to only that which is likely to bring into effect the communicator's intent — its legal prohibition should. Waldron recognises a "likely to" requirement as reflective of the importance of hate speech's effect.[145] This explains its inclusion in foreign hate speech prohibitions. In Canada, s 319(1) requires the incitement of hatred to be "likely to lead to a breach of the peace".[146] In the United Kingdom, the communicator must either intend to stir up racial hatred or "having regard to all the circumstances racial hatred is likely to be stirred up".[147] Although I believe this provision would be improved if both elements were required, the inclusion of "likely to" demonstrates the effect of the communication is considered an important element by foreign jurisdictions. Moreover, it is a requirement for a cause of action in defamation, a less serious class of prohibited communications. In New Zealand, the defamatory statement must be likely to "cause the person to be exposed to hatred,

---

[142] The Royal Commission intended this to be the result, as they believed if a defendant had made an abusive or insulting publication with an intent to incite hatred, whether the intention was likely to result was an unnecessary requirement: Royal Commission, above n 14, at [46].

[143] Parekh, above n 52, at 41.

[144] At 41 (emphasis added).

[145] Waldron, above n 55, at 35.

[146] Criminal Code RSC 1985 c C-46, s 319(1).

[147] Public Order Act (UK), s 18(1)(a).

contempt or ridicule”.[148] If a notional link between the conduct and the outcome is required for civil defamation, a criminal hate speech provision which is intended to have a high liability threshold,[149] and carries a possible three-year sentence,[150] should require the same.

Contrary to the proposed provision, foreign provisions that only require intention typically have a narrower application and/or impose a higher liability threshold. In Canada, s 319(2) of the Criminal Code requires the defendant “wilfully promote[d] hatred”.[151] Unlike subs 1, this provision is subject to several exceptions: for example, if the statements were relevant to any subject of public interest; or if the speaker was, in good faith, attempting to establish an argument on a religious subject.[152] Canadian courts have applied an equally narrow interpretation of the offence: “wilfully promotes” requires that “the hate-monger must intend or foresee as substantially certain a direct and active stimulation of hatred against an identifiable group”.[153] This is also manifest in the United Kingdom. While s 18 of the Public Order Act targets racial hatred, s 29B captures religious hatred and only requires that the defendant, when communicating threatening words or behaviour, intended to stir up religious hatred.[154] However, s 29J of the Racial and Religious Hatred Act exempts from s 29B of the Public Order Act statements of critique, insult or abuse against religions or the beliefs or practices of their adherents.[155] It also protects statements that urge adherents of a particular religion to cease practising their religion. The application of s 29B is therefore extremely narrow.[156]

In light of these foreign provisions, I argue it is short-sighted to have a generalised hate speech provision as only requiring intention. At the very least, and taking guidance from

---

148 Ian McKay *Laws of New Zealand* Defamation (online ed) at [10], citing *Parmiter v Coupland* (1840) 6 M & W 105 (Ex) at 108.

149 Royal Commission, above n 14, at [42].

150 Ministry of Justice, above n 10, at 32.

151 Criminal Code RSC 1985 c C-46, s 319(2).

152 Criminal Code RSC 1985 c C-46, s 319(3).

153 *R v Topham*, above n 66, at [60], citing *R v Keegstra*, above n 75, at 776–777.

154 Section 29B(1). Note that s 29B also captures hatred on the grounds of sexual orientation.

155 Racial and Religious Hatred Act 2006 (UK), s 29J.

156 Natalie Alkiviadou “Regulating Hate Speech in the EU” in Stavros Assimakopoulos, Fabienne Baider and Sharon Millar (eds) *Online Hate Speech in the European Union: A Discourse-Analytic Perspective* (Springer International Publishing, Cham, 2017) 6 at 10.

foreign legislation, if the "likely to" requirement is to remain omitted, then certain exceptions and defences should be inserted to narrow its application. In fact, such exemptions are desirable even when divorced from the "likely to" requirement, as it would aid in alleviating concerns regarding encroachment on free speech. Alexander Brown describes how these are a typical feature of the "incitement of hatred" cluster of hate speech laws and exempt such communications as "statements published or broadcast for the purposes of public debate, artistic expression, journalistic reporting, or editorial commentary".[157] Such provisos would aid in narrowing the offence's application, resulting in a minimum infringement upon freedom of expression.

Ideally then, the proposed provision should be amended to include a "likely to" requirement in addition to intention, as well as exemptions like those present overseas. The core justification for including a "likely to" requirement boils down to the fact that it is the likely effect of hate speech that distinguishes it as an offence and, therefore, that effect should be an element of determining culpability. Thus, I disagree with the Commission's argument that this element has "little or no bearing on whether the conduct is sufficiently culpable to justify [a charge]".[158] This is especially so in the context of the added intentions of "normalise" and "maintain". Prima facie, the passivity of these verbs means they will be easier to prove than incitement, a point which has raised concerns of unjustifiably lowering the liability threshold.[159] Take a scenario where a defendant's conduct is perhaps the lowest form of culpability under the provision: intentionally maintaining hatred through insulting communications. That their intention is likely to result is a strong indicator as to whether this culpability is sufficient to justify a charge.[160] Considering the status of freedom of expression, the objective "likely to" requirement is a necessary safeguard against unwarranted charges, charges which would show the provision infringes upon the freedom more than is reasonably necessary, thus failing to meet the "demonstrably justified" standard.[161]

---

157 Brown, above n 51, at 24.

158 Royal Commission of Inquiry, above n 14, at [46].

159 Epati, above n 121, at [4.6].

160 A point also raised by the New Zealand Law Society: Epati, above n 121, at [4.6].

161 New Zealand Bill of Rights Act, s 5.

VI Conclusion

The proposal to replace s 131 of the HRA with a new provision in the Crimes Act has strong justifications. Section 131 is inadequate in several areas, its lack of application to electronic communications being the most blatant. Hate speech produces a very real harm in New Zealand, increasingly through its online presence. I agree it should be addressed by the criminal law, and a provision in the Crimes Act would reflect the seriousness of its harm. Not only does hate speech create a culture whereby the vilification, discrimination and violence against a protected group are legitimised, it also hampers freedom of expression by making it highly undesirable for members of the targeted group to speak back. Thus, the debate is not as simple as freedom of expression versus the harm of hate speech. Freedom of expression, however, includes the right to communicate hate speech. The issue of limiting this is complicated by the inviolable status the freedom holds in Western democracy. There are good reasons for this status — the marketplace of ideas, for example. Its protection by s 14 of NZBORA requires that any limitation upon it must be "demonstrably justified".[162] This means any New Zealand hate speech law must be a minimal infringement upon the freedom.[163] To achieve this, the provision must be carefully and clearly drafted to ensure a narrow application that does not result in unjustified charges.

The Royal Commission's goal of creating judicial certainty with the provision is commendable, however, more definitional clarity is needed to achieve this. Replacing the terminology of s 131 with "hatred" will not act as a silver bullet to defeat imprecise interpretations, as evidenced by *Wall*. "Hatred" needs a clear definition that distinguishes it from the emotional expression of hate. The removal of the "likely to" requirement is the most concerning aspect of the proposal. Both this and intention should be required, as it is by s 131 of the HRA. Moreover, provisos should be inserted to exempt certain communications, for example, those made in a public debate. This would bring New Zealand in line with foreign jurisdictions and provide a lesser infringement upon freedom of expression, thus making it more likely to be considered demonstrably justified under NZBORA.

---

[162] New Zealand Bill of Rights Act, s 5.

[163] *R v Hansen*, above n 114, at [104] per Tipping J.

If hate speech legislation is to be done properly, great care needs to be taken in its drafting, more so than the status quo for other criminal provisions. The unique status of prohibited speech requires this — communication is far more ambiguous than physical action. This care is arguably missing from the proposal, perhaps a symptom of it being one of 44 recommendations that the Government agreed to implement.[164] The proposal, therefore, is a promising first step, but should not be the decisive answer.

Word Count

The text of this paper (excluding the table of contents, abstract, footnotes and bibliography) comprises exactly 8,161 words.

---

[164] See also Bromell, above n 54, at 6, noting the rushed nature of the proposal is also manifest in the Government's failure to carry out their promise of a "robust public discussion from all quarters".

VII Bibliography

A Cases

1 New Zealand

King-Ansell v Police [1979] 2 NZLR 513 (CA).

Proceedings Commissioner v Archer (1996) 3 HRNZ 123 (CRT).

R v Hansen [2007] NZSC 7, [2007] 3 NZLR 1.

Wall v Fairfax [2018] NZHC 104. [2018] 2 NZLR 471.

Young v Cassells (1914) 33 NZLR 852 (HC).

2 Australia

Sunol v Collier (No 2) [2012] NSWCA 44, (2012) 289 ALR 128.

3 Canada

R v Keegstra [1990] 3 SCR 697.

R v Sears [2019] ONCJ 104.

R v Topham [2017] BCSC 259.

Saskatchewan Human Rights Commission v Whatcott 2013 SCC 11, [2013] 1 SCR 467.

4 United Kingdom

Parmiter v Coupland (1840) 6 M & W 105 (Ex).

5 United States

Abrams v United States 250 US 616 (1919).

B Legislation

1 New Zealand

Crimes Act 1961.

Human Rights Act 1993.

New Zealand Bill of Rights Act 1990.

Sentencing Act 2002.

2 Canada

Criminal Code RSC 1985.

3 Germany

Strafgesetzbuch [German Criminal Code] 1872.

4 United Kingdom

Public Order Act 1986.

Racial and Religious Hatred Act 2006.

C Treaties

International Convention on the Elimination of All Forms of Racial Discrimination 660 UNTS 195 (opened for signature 21 December 1965, entered into force 4 January 1969).

D Books and Chapters in Books

Alexander Brown *Hate Speech Law: A Philosophical Examination* (Routledge, New York and Oxford, 2015).

Bhikhu Parekh "Is There a Case for Banning Hate Speech?" in Michael Herz and Peter Molnar (eds) *The Content and Context of Hate Speech: Rethinking Regulation and Responses* (Cambridge University Press, New York, 2012) 37.

Caroline West "Words that Silence? Freedom of Expression and Racist Hate Speech" in Ishani Maitra and Mary Kate McGowan (eds) *Speech and Harm: Controversies over Free Speech* (Oxford University Press, Oxford, 2012) 222.

David Bromell *Ethical Competencies for Public Leadership: Pluralist Politics in Practice* (Springer International Publishing, Cham, 2019).

Douglas Fraleigh and Joseph Tuman *Freedom of Expression in the Marketplace of Ideas* (SAGE, London, 2011).

Edwin Baker "Hate Speech" in Michael Herz and Peter Molnar (eds) *The Content and Context of Hate Speech: Rethinking Regulation and Responses* (Cambridge University Press, New York, 2012) 57.

Ian McKay *Laws of New Zealand* Defamation (online ed).

Jeremy Waldron *The Harm in Hate Speech* (Harvard University Press, 2012).

Joel Feinberg *Rights, Justice, and the Bounds of Liberty: Essays in Social Philosophy* (Princeton University Press, Princeton, 1980).

Joel Feinberg *The Moral Limits of The Criminal Law Volume 1: Harm to Others* (Oxford University Press, New York, 1987).

John Stuart Mill *On Liberty (1859)* (Batoche Books, Kitchener, 2001).

Katharine Gelber and Adrienne Stone (eds) *Hate Speech and Freedom of Speech in Australia* (The Federation Press, Sydney, 2007).

Katharine Gelber *Speaking Back: The Free Speech Versus Hate Speech Debate* (John Benjamins Publishing Company, Amsterdam and Philadelphia, 2002).

Michael Herz and Peter Molnar "Introduction" in *The Content and Context of Hate Speech: Rethinking Regulation and Responses* (Cambridge University Press, New York, 2012) 1.

Natalie Alkiviadou "Regulating Hate Speech in the EU" in Stavros Assimakopoulos, Fabienne Baider and Sharon Millar (eds) *Online Hate Speech in the European Union: A Discourse-Analytic Perspective* (Springer International Publishing, Cham, 2017) 6.

Peter Molnar "Responding to 'Hate Speech' with Art, Education, and the Imminent Danger Test" in Michael Herz and Peter Molnar (eds) *The Content and Context of Hate Speech: Rethinking Regulation and Responses* (Cambridge University Press, New York, 2012) 183.

E Journal Articles

Anne Flahvin "Can Legislation Prohibiting Hate Speech be Justified in Light of Free

Speech Principles?" (1995) 18 UNSWLJ 327.

Colin Bird "The Possibility of Self-Government" (2000) 94(3) APSR 563.

Joe Dryden "Protecting Diverse Thought in the Free Marketplace of Ideas: Conservatism and Free Speech in Higher Education" (2018) 23(1) Tex Rev L & Pol 229.

John Ip "Debating New Zealand's Hate Crime Legislation: Theory and Practice" (2005) 21 NZULR 575.

Juliet Moses "Hate Speech: Competing Rights to Freedom of Expression" (1996) 8 AULR 185.

Matthew Williams, Pete Burnap, Amir Javed, Han Liu and Sefa Ozalp "Hate in the Machine: Anti-Black and Anti-Muslim Social Media Posts as Predictors of Offline Racially and Religiously Aggravated Crime" (2020) 60(1) Br J Criminol 93.

Melina Bell "John Stuart Mill's Harm Principle and Free Speech: Expanding the Notion of Harm" (2021) 33(2) Utilitas 162 at 168.

Michael Conklin "The Overlooked Benefits of 'Hate Speech': Not Just the Lesser of Two Evils" (2020) 60 S Tex Law Rev 687.

Stephen Guest "Why Hate has no Place in the Criminal Law" [2021] NZLJ 371.

F Parliamentary and Government Materials

(8 December 2020) 749 NZPD (Royal Commission of Inquiry into the Terrorist Attack on Christchurch Masjidain on 15 March 2019 – Report, Rt Hon Jacinda Ardern (Prime Minister)).

Department of the Prime Minister and Cabinet *Report on Community Hui Held in Response to the Royal Commission into the Terrorist Attack on Christchurch Mosques on 15 March 2019* (March 2021).

Ministry of Justice *Proposals Against Incitement of Hatred and Discrimination* (June 2021).

Royal Commission of Inquiry *Hate speech and hate crime related legislation* (November 2020).

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

"The Christchurch mosque attacks: how Parliament responded" (11 March 2021) New Zealand Parliament <www.parliament.nz>.

G Reports

David Bromell *'Hate Speech': Defining the Problem and Some Key Terms* (Institute for Governance and Policy Studies, Victoria University of Wellington, Working Paper 21/03, March 2021).

H Dissertations

Elizabeth Macpherson "Regulating Hate Speech in New Zealand" (LLB (Hons) Research Paper, Victoria University of Wellington, 2003).

I Internet Resources

Benn Bathgate "Man who went 'down the rabbit hole' sentenced for video that incited hate against Māori" (22 April 2022) Stuff <stuff.co.nz>.

"Christchurch mosque terror attacks day 8: What you need to know" (22 March 2019) Radio New Zealand <rnz.co.nz>.

Edgar Pacheco and Neil Melhuish "Measuring Trends in Online Hate Speech Victimisation and Exposure, and Attitudes in New Zealand" (December 2019) Netsafe <www.netsafe.org.nz>.

"Preliminary Submission on the Proposed Hate Speech Legislation" The Federation of

Islamic Associations of New Zealand (Inc.) <www.fianz.com>.

J Other Resources

Letter from Tiani Epati (President of the New Zealand Law Society) to the Ministry of Justice regarding the proposals against incitement of hatred and discrimination (5 August 2021).