Threats From the Fringe: *Understanding and Solving Fringe Social Media Sites' Threats and Incitements to Violence Against New Zealand's MPs.*

Meredith Ross-James
Parliamentary Intern, New Zealand Parliament.

Abstract: *Online hate and harassment against MPs is rising, not just in New Zealand, but internationally. Given the highly public nature of the internet, many threats are being made on public forums for the world to see. Recent events of mass online violent harassment of MPs and other political figures in New Zealand have shown that our legislation is not fit for purpose. But what is that purpose? And how might we go about meeting the needs that contemporary hate groups create? This paper examines violence-speech and death threats publicly made against New Zealand MPs and political figures on the fringe social media site Gab, uncovering the ways New Zealand's legislation has thus far failed to address this dangerous behaviour, and how legislation may be shaped to address it. After an analysis of top posts in New Zealand messageboards on Gab, it was uncovered that a significant majority of threats were made impersonally and indirectly- that is, most threats were expressing a desire for or approval of violence against an unspecified group. The second largest set of threats were made against specified individuals, but were incitory in nature, and did not implicate the poster in the violence. This, alongside a distinct lack of incitement to violence laws, shows a significant need in New Zealand for anti-violence incitement laws similar to those passed in Australia. While the first half of this paper investigates these death threats, the second half takes a comparative approach to incitement to violence laws in Australia and considers how these could be implemented in New Zealand's legislation for the benefit of all political staff and other vulnerable communities.*

Introduction

Recent research has shown that hate against and harassment of Members of Parliament (MPs) in Aotearoa New Zealand is rising. Especially after COVID-19 lockdowns, the 2022 occupation of parliament, and the 2023 election, MPs are subject to higher levels of hate than ever before.[1] Unmoderated, fringe social media communities represent a significant site wherein violent sentiment against 'othered' individuals is facilitated and encouraged, causing both psychological and real-world harm against targeted groups.[2,3] The violence (and threats thereof) incurred by these communities against political officials also degrades the quality of democracy by discouraging people with marginalised identities from running for office. [4,5] Fearful journalists can not effectively report on political news when they're subject to death threats for doing their jobs,[6, 7] and this threat too is becoming more pronounced in Aotearoa.[8] Not only is online hate and harassment against politicians rising, but new initiatives to regulate online and real-world political violence in Aotearoa have ceased. Aotearoa's project to establish an agency regulating social media platforms was disestablished in May 2024, and recommendations of the Royal Commission inquiry into the 2019 Christchurch terrorist attack were abandoned soon after.[9,10,11] Online violence facilitated in fringe social media communities and the real-world violence and harm resulting therefrom has ripple effects throughout the whole democratic structure when performed against and political officials in particular. In the interests of maintaining the health, diversity, and stability of our democratic structures, then, we can not ignore violent fringe and extremist social media communities.

---

[1] Susanna Every-Palmer, Oliver Hansby, and Justin Barry-Walsh, 'Stalking, Harassment, Gendered Abuse, and Violence Towards Politicians in the COVID-19 Pandemic and Recovery Era'. *Front. Psychiatry* 15 (1) 2024, pp. 5, 8.
[2] Arne Dreißigacke, Philipp Müller, Anna Isenhardt, and Jonas Schemmel, 'Online Hate Speech Victimisation: Consequences for Victims' Feelings of Insecurity'. *Crime Science* 13 (4) 2024, p. 7.
[3] Human Rights Campaign, 'New Human Rights Campaign Foundation Report: Online Hate & Real World Violence Are Inextricably Linked'. *Human Rights Campaign,* December 13 2022.
[4] Every-Palmer, Hansby, and Barry-Walsh, Stalking, Harassment, Gendered Abuse, and Violence, p, 8.
[5] Maya Oppenheim, 'General election: Women MPs Standing Down Hver 'Horrific Abuse', Campaigners Warn'. *The Independent,* October 31 2019.
[6] National Union of Journalists, 'Four in Five Journalists Have Experienced Threats and Violence at Work'. *National Union of Journalists*, November 3 2021.
[7] Silvio Waistbord, 'Mob Censorship: Online Harassment of US Journalists in Times of Digital Hate and Populism'. *Digital Journalism* 8 (1) 2020, pp.2037-38.
[8] Susan Fountaine and Cathy Strong, 'An Intersectional Analysis of Aotearoa New Zealand Journalists' Online and Offline Experiences of Abuse, Threats and Violence'. *Journalism Studies* 25 (1) 2023, 168.
[9] Tom Pullar-Strecker, 'Internal Affairs Scraps Ambitious Plan to Clean up the Internet'. *The Post*, May 10 2024.
[10] Tom Pullar-Strecker, 'Internal Affairs Sets Out Big Plans to Regulate Harmful Content on Social Media'. *Stuff*, June 1 2023.
[11] Radio New Zealand, 'Christchurch terror attacks: Government's binning of recommendations 'shameful', Muslim leader says'. *Radio New Zealand,* August 2 2024.

An egregious example of a community of this sort that has directly impacted MP and staff safety in Aotearoa is the ‘Nuremberg 2.0’ website that existed from October 2021 until April 2022.[12] Users posted which government officials, ministry workers, journalists, or activists they believed deserved to be charged in the ‘Nuremberg 2.0’ trials, what violent punishment they ‘deserved’, and for what reason. This sort of group communication- the publishing of public ‘hit lists’- is a common, low-cost high-yield threat device employed by communities of hate.[13] Unsurprisingly, then, the public online community of ‘Nuremberg 2.0’ was full of vitriolic calls for the killing of political officials. Posts like “All those involved and complicit need to be rounded up, tried for those crimes, and given no less than the death penalty” were prolific.[14]

Those subject to public death threats and public harassment of this sort struggle to stop this behaviour. One New Zealand academic who researches and publishes on the extreme-right has been subject to public threats calling for him to be hanged,[15,16] and these public threats have led to the harasser’s social media community joining the harassment campaign against him.[17] A prolific science communicator, too, was named on the ‘Nuremberg 2.0’ site, where users publicly called for her to be killed.[18] Neither of these high-profile academics were able to seek justice for the public threats made against them, because laws are not fit for the contemporary hate and harassment landscape. The police, the chief censor, and the Domain Name Commission (who control the .nz domain the website used), all lamented that they were unable to take action to shut down the ‘Nuremberg 2.0’ website because the legislation did not exist to allow them to do so.[19] While threats do not always indicate that someone will commit violence, and people who commit violence do not always make threats beforehand, social media platforms hosting violent communities of hate like Nuemberg 2.0, Gab, 4chan, and 8chan radicalise their users, facilitate anger that motivates action, and create social endorsement of such action.[20] A thriving and unimpeded ecosystem of hateful social media

---

[12] Toby Manhire, ‘Inaction on NZ ‘Nuremberg’ Site Sparks Calls for Overhaul of System ‘Not Fit for Purpose’’. *The Spinoff*, April 4 2022.
[13] Enrique Eguren, ‘Understanding Death Threats Against Human Rights Defenders’. *Protection International,* 2021, pp. 2-3.
[14] Manhire, ‘Inaction on NZ ‘Nuremberg’ Site’.
[15] Martyn Bradbury, ‘Siouxsie Wiles Gets a Pittance and Byron Doesn’t get Protection – All Public Academics and Researchers Will Suffer’. *The Daily Blog*, July 13 2024.
[16] Nadine Roberts, ‘What it Took to Stop Harassment From a White Supremacist’. *Stuff*, September 29 2023.
[17] Roberts, ‘What it Took to Stop Harassment’.
[18] Emma Stanford, ''Wrong on Many Levels': Disinformation Researcher Criticises University of Auckland's 'Silencing' of Siouxsie Wiles’. *Radio New Zealand,* November 14 2023.
[19] Manhire, ‘Inaction on NZ ‘Nuremberg’ Site’.
[20] Brian Ballsun-Stanton, Lise Waldek, Julian Droogan, Debra Smith, Muhammad Iqbal, and Mario Peucker, ‘Mapping Networks and Narratives of Online Right-Wing Extremists in New South Wales: Final Report’. *Macquarie University, Department of Security Studies and Criminology* 2020, pp. 30, 32.

communities nurtures these sentiments and intentions, and for that reason, we must pay close attention to them.

Fringe, violent social media communities and the threats of violence against political staff that result therefrom demands a response in order to protect the quality and diversity of our democracy. Giving groups power to take action on these issues, though, requires legislative change, which itself requires a thorough understanding of the exact shape of the problem. This paper, then, begins literature review, locating the problem of online violence-speech against political officials in international contexts and understanding similar studies' findings. The review will both provide a hypothesis and a direction for a study design, which will follow. The forthcoming study analyses the nature of the communications threatening and inciting violence against MPs in New Zealand, with a vision to shape solutions tailor-made to deal with the actual nature of communications within these violent fringe social media communities.

Literature Review

What does the literature indicate about violence-speech and online death threats against politicians? In 2021, Peter Eisler, Jason Szep, Linda So, and Sam Hart conducted research for Reuters into the form of death threats against political officials and election workers following Donald Trump's 2020 election loss.[21] They collected 850 threatening messages through interviews, online post collection, and public records requests. Only 13% were 'prosecutable'. In the remaining 87% of threats, researchers say, the "harassers call for violence without threatening to act themselves".[22] Even utterances like "You and your family will be killed very slowly" were 'unprosecutable'.[23] The researchers indicate that even though threats may be directed towards an individual, speakers often indicate no *personal* intent. The lack of personal intent does not, of course, reduce the psychological harm incurred by the victim. Their research focuses primarily on the legality of the threats made against officials- that being their 'true intent'. In the United States, a death threat is illegal when it constitutes a "true threat", rather than "political hyperbole" or "emotionally charged rhetoric".[24] Reuters' legal scholars sorted Eisler et. al's data into *likely legal* or *potentially illegal* on that basis. In the former category were utterances like "patriots are coming for

---

[21] Peter Eisler, Jason Szep, Linda So and Sam Hart, 'Anatomy of a Death Threat'. *Reuters* December 30, 2021.
[22] Eisler, Szep, So, and Hart, 'Anatomy of a Death Threat'.
[23] Eisler, Szep, So, and Hart, 'Anatomy of a Death Threat'.
[24] Constitution Annotated. 'Amdt 1.7.5.6 True Threats'. Accessed at: https://constitution.congress.gov/browse/essay/amdt1-7-5-6/ALDE_00013807/.

you", "This woman should be strung up in the goddamn state capitol", and "if your children can't be tried for treason like you, then I pray that your children get cancer. And die a slow, miserable death. And you have to watch". In the latter category were utterances such as "let's burn her house down and kill her family", "they will be hung for treason", and "prepare for the gallows".[25] The lexical difference between many of these examples is slight. While 'this woman should be [hung] at the state capitol' is likely protected, 'they will be hung for treason' likely is not. Indeed, it seems that the former may be more indirect, using the indefinite modal verb 'should' being perhaps more permissible than the definite modal verb 'will'; one indicating a desire and another a sense of certainty and intent. This is mirrored in the 'potentially illegal' utterances '*let's* burn her house down' and the imperative sentence 'prepare for the gallows'.[26]

Bjørgo et. al (2022) consider violent threats made against Norwegian politicians because, they say, when threats occur against a background of violent attacks on politicians, threats become acts of political violence *in themselves*.[27] Asking Norwegian politicians and party youth-wing members about their experiences of online and in-person harassment, they found that while online threats were less common than online harassment, *indirect* threats were received far more frequently than *direct* threats (the former being experienced by 40% of parliamentarians in 2021, and the latter by 28%).[28] Threats are measured as threats *to harm* the individual or those close to them,[29] but the difference between direct and indirect threats was not specified.[30] If we take that 'direct' and 'indirect' track broadly onto the 'prosecutable' and 'not prosecutable' distinction outlined in Eisler et. al, we see again that indirect, non-prosecutable threats against politicians occur at a significantly higher rate than the 'prosecutable' direct threats. This disparity, of course, does not mean the victims of such threats experience lower levels of psychological damage as a result.[31]

The 2022 Threats and Harassment Against Local Officials Dataset, created by a coalition of groups including Princeton's Bridging Divides Initiative and the Anti-Defamation League, analysed approximately 3,000 unique incidents of threats and harassment against United

---

[25] Eisler, Szep, So, and Hart, 'Anatomy of a Death Threat'.
[26] Eisler, Szep, So, and Hart, 'Anatomy of a Death Threat'.
[27] Tore Bjørgo, Anders Ravik Jupskås, Gunnar Thomassen, and Jon Strype, 'Patterns and Consequences of Threats Towards Politicians: Results from Surveys of National and Local Politicians in Norway'. *Perspectives on Terrorism* 16 (6) 2022, p. 101.
[28] Bjørgo, Jupskås, Thomassen, and Strype, 'Patterns and Consequences of Threats,' pp. 105-06.
[29] Bjørgo, Jupskås, Thomassen, and Strype, 'Patterns and Consequences of Threats,' p. 103
[30] Bjørgo, Jupskås, Thomassen, and Strype, 'Patterns and Consequences of Threats,' p. 105
[31] Bjørgo, Jupskås, Thomassen, and Strype, 'Patterns and Consequences of Threats,' p. 113.

States local (not federal or state) officials.[32] They define threats as utterances communicating intentions to inflict pain or hostile action on an individual because of their role as an official, that would reasonably cause the victim to fear for their or their family's safety. Considering this, the dataset counts both legal and illegal threats. This is contrasted against 'harassment', which intends to intimidate, threaten, or terrorise the victim.[33] The group utilised human coding, each incident being coded by two researchers after it was approved by two other researchers for eligibility.[34] While the researchers did not divide threats into legal and illegal threats, they did sort them by topic. Of the threats identified, 34% were threats of death or gun violence,[35] while 18% were threats to perform multiple acts of violence, and 14% were ambiguous about the nature of the violence. Other specified threats were in the minority.[36]

We must pay attention to death threats made *within* communities of hate, as well as threats made directly to victims, as research shows threats made within this context amplify the group members' desire and endorsement of violent acts. Particularly, research suggests group posts with motivational 'collective action' frames (that is, posts that call for or acknowledge a *need* for group action to address a grievance) prepare people for offline mobilisation. These kinds of posts promote political violence as in-group endorsed, morally justified, legitimate ways to pursue a 'common cause'.[37] When members of the in-group repeatedly identify existential threats, members are more likely to endorse violence. Bailard et. al calls this phenomenon 'moral convergence'.[38] In their study of over 500,000 posts on 'Proud Boys' Telegram channels, Bailard et. al found that increases in the number of posts with a motivational frame correlated with increased instances of Proud Boys' violence offline,[39] while posts with a prognostic frame- a frame *specifying violent solutions* to group problems- did not.[40] This sentiment is repeated in Jess Berentson-Shaw and Marianne Elliott's paper *Online Hate and Offline Harm*. They describe the phenomenon of the *false consensus effect*, where those with extreme views believe their views are more widely

---

[32] Joel Day, Aleena Khan, and Michael Loadenthal, 'Threats and Harassment Against Local Officials Dataset'. *Bridging Divides Initiative* 2022, p. 4.
[33] Day, Khan, and Loadenthal, 'Threats and Harassment Against Local Officials,' p. 11.
[34] Day, Khan, and Loadenthal, 'Threats and Harassment Against Local Officials,' p. 13.
[35] Day, Khan, and Loadenthal, 'Threats and Harassment Against Local Officials,' p. 5.
[36] Day, Khan, and Loadenthal, 'Threats and Harassment Against Local Officials,' p. 22.
[37] Catie Bailard, Rebekah Tromble, Wei Zhong, Frederico Bianchi, Pedram Hosseini, and David Broniatowski, '"Keep Your Heads Held High Boys!": Examining the Relationship between the Proud Boys' Online Discourse and Offline Activities'. *American Political Science Review* 2024, pp. 2, 14.
[38] Bailard, Tromble, Zhong, Bianchi, Hosseini, and Broniatowski, '"Keep Your Heads Held High Boys!," p. 14.
[39] Bailard, Tromble, Zhong, Bianchi, Hosseini, and Broniatowski, '"Keep Your Heads Held High Boys!," pp. 9, 12.
[40] Bailard, Tromble, Zhong, Bianchi, Hosseini, and Broniatowski, '"Keep Your Heads Held High Boys!," p. 15.

shared than they actually are.[41] If one spends time in a violent social media community that openly discusses the desire or necessity for violence, one's belief that political violence is warranted, acceptable, or even *necessary* becomes falsely normalised. Above all, it is clear that even when violence-speech and death threats are not received by victims themselves, these kinds of threats encourage, normalise, and legitimise the use of violence against officials. This has been acknowledged by counterterrorism scholarship, which points out that 'Lone Wolf' actors are never *truly* alone, as they often have online communities of hate that radicalised, educated, and incited them to commit their acts.[42]

Given the scholarship coming out of other states, we can predict that a majority of posts endorsing or calling for violence against MPs made in violent fringe social media communities are 'non-prosecutable'. At first glance, many seem to be. While the site is inaccessible now, journalist Toby Manhire archived some '*Nuremberg 2.0'* posts in his article on the site. "Deserves the rope to be hung till death and nothing more or less will be sufficient" says one commenter. "[T]he Govt. members should face the death penalty" says another.[43] The comment "All those involved and complicit need to be rounded up, tried for these crimes and given no less than the death penalty" mirrors Eisler et. al's observation that many 'non-prosecutable' threats involve calls for the use of legal means.[44,45] Just as the vast majority of threats examined by Eisler et. al and Day, Khan, and Loadenthal in the Threats Against Local Officials Dataset were non-prosecutable; so too was Nuremberg 2.0 and the threats posted thereupon immune to any legal recourse by Aotearoa's police, security, and cybersecurity officials.[46]

Given the outcomes of the studies surveyed above, it is hypothesised that we will see two kinds of threats emerge in Aotearoa's violent fringe social media communities. First, there will be the 'direct' or 'prosecutable' threats: threats made against a distinct individual that indicate personal intent to commit the act. Second, we can expect to see 'indirect' or 'non prosecutable' threats: threats made to a vague class of individuals that endorse or incite some act of violence. These indicate four distinct states a threat can be in: it can be direct and personal- threatening an individual with personal intent, or it can be direct and impersonal- threatening an individual with impersonal, incitory intent. Threats may also be

---

[41] Jess Berentson-Shaw and Marianne Elliott, 'Online Hate and Offline Harm'. *The Workshop* 2019, pp. 14-15.
[42] Royal Commission of Inquiry into the Terrorist Attack On Christchurch Mosques on 15 March 2019, 'Part 2: Context'. *Royal Commission of Inquiry into the Terrorist Attack On Christchurch Mosques on 15 March 2019* 1 (1) 2020, p. 109.
[43] Manhire, 'Inaction on NZ 'Nuremberg' Site'.
[44] Manhire, 'Inaction on NZ 'Nuremberg' Site'.
[45] Eisler, Szep, So, and Hart, 'Anatomy of a Death Threat'.
[46] Manhire, 'Inaction on NZ 'Nuremberg' Site'.

impersonal and direct- made against a group of unspecified individuals with personal intent, or impersonal and indirect- made against a group of unspecified individuals with impersonal, incitory intent. As seen before, these distinctions can be understood through a lexical and semantic analysis of posts: a consideration of their modal verbs, singular or plural personal pronouns, and more. Scholars have endorsed discourse analysis as an effective method for analysing language's justifying effect on violence. Through 'microdetails', people morally situate themselves in regards to the actions they endorse. They allow speakers and groups to claim or negate responsibility, establish collective responsibility, or implicate others in their described acts.[47] Finally, we can expect that the majority of these threats will be in the 'indirect' 'non prosecutable' category. That is, any threat that is not personal and direct. Should this be the case, which the case of the Nuremberg 2.0 website indicates it may be, it will demonstrate a significant need to tailor response strategies to communities that create violent threats in this form. With background and expectations established we may move onto the design of this study.

Method

To understand the makeup of violence-speech and threats against New Zealand's politicians and political officials, an analysis was performed on public violence-threats made against these groups on extremist fringe social media site Gab. Gab was chosen for a number of reasons. First, because it hosts the most New Zealand extremist activity of any non-mainstream social media platform.[48] Second, because of its public accessibility; one does not need an account to see posts on Gab. Third, because of its relaxed speech rules. Gab is (in)famous for hosting and radicalising terrorists,[49] and their founder has forthrightly refused to implement speech restrictions on the platform.[50] Fourth, because Gab hosts topic-specific boards, which makes finding discourse related to a particular country very easy. 'Most popular' posts (sorted by highest number of comments and reactions) on 'New Zealand Politics' groups, or groups similar thereto, were analysed to capture sentiment towards Aotearoa's politicians specifically. Many studies of this type use large API 'scraping' or 'snowballing' models to acquire content for analysis and computer methods for coding,[51]

---

[47] Robin Conley Riner, 'Language and Violence'. *Oxford Research Encyclopedia of Anthropology* 2023, p. 12.
[48] Milo Comerford, Jakob Guhl and Carl Miller, 'Understanding the New Zealand Online Extremist Ecosystem'. *Institute for Strategic Dialogue* 2021, p. 10.
[49] Rita Katz, ' Inside the Online Cesspool of Anti-Semitism That Housed Robert Bowers'. *Politico Magazine*, October 29 2018.
[50] Christopher St.Aubin and Galen Stocking, 'Key Facts About Gab'. *Pew Research Center,* January 24 2023.
[51] Ballsun-Stanton, Waldek, Droogan, Smith, Iqbal, and Peucker, 'Mapping Networks and Narratives,' pp. 11, 14.

but due to technological restrictions and the semantic specificity of violence-talk and threats,[52] manual data-gathering was performed. When threats of violence were noticed, they were captured, archived, and manually coded. Given the observational nature of this study, no independent variable is manipulated, but the dependent variable for quantitative measurement was the category into which each instance of violence-speech was coded: direct or indirect, and personal or impersonal.

For the purposes of this specific study, any post endorsing violence against any current or past Member of Parliament, as well as parliamentary or electorate staff was included for analysis, whether or not that violence was indicated to be in pursuit of a political goal, and whether or not the speech would reasonably cause the victim to dear for their or their loved ones' safety. This includes both explicit 'death threats' *and* endorsements of violence that are usually protected. Terms like 'violence-speech' 'death threat' and 'violent threat' are generally interchangeable in this very broad context, but they are united in their communication of intention, encouragement, or positive recognition of hostile action taken against some individual.[53] A definition of violence-speech for the purposes of this study must be broad by necessity so the broad milieu of violence-speech can be effectively understood with a coding framework that operationalises speech-features for regulation or response.

Criteria for each coding category are as follows: First, a *personal* threat indicates personal intention, whereas an *impersonal* threat indicates incitement, personal endorsement, or agreement with acts of violence. "I will x" or "I want to give her x" indicates personal intent primarily with the personal pronoun "I". Utterances like "She deserves x", "someone ought to x her" or utterances of approval to instances of violence-speech like "true that!" are *impersonal*, as they do not indicate any intent by the speaker to act, but personal endorsement or acceptance of violence. The second measure employed is whether some threat is made *directly* or *indirectly*. A *direct* threat is made against a specified individual or a group of specified individuals. For example, "she deserves x" is a direct threat as the singular pronoun "she" indicates a specific individual, specified by the context of surrounding utterances. Direct threats also include utterances like "John Doe deserves x", "I will x John Doe", or "every member of Party y must be x-ed". Finally, an *indirect* threat captures threats made against an unspecified group or class of individuals. An indirect threat may look like "people like that deserve x" or "I will x any politician who denies our freedoms". 4 possible coding options were available for threats encountered in this study, then: Impersonal Direct

[52] Riner, 'Language and Violence,' 9.
[53] Day, Khan, and Loadenthal, 'Threats and Harassment Against Local Officials,' p. 11.

(ID) “John Doe deserves to be hanged”, Personal Direct (PD) “I will hang John Doe”, Impersonal Indirect (II) “people who deny our rights deserve to get hanged”, and Personal indirect (PI) “I will hang anyone who denies us our rights”. After data collection, the prevalence of each kind of threat will indicate directions for online violence prevention strategies to take.

Results

The Gab boards *New Zealand Politics* (2,600 members) and *New Zealand Gab* (1,500 members) were observed. On these boards, the top 10 most popular posts (measured by number of reactions and comments) were selected for analysis. A total of 810 posts were analysed. This includes 20 ‘parent’ posts and 790 comments. Of the posts, 5.1% (42), or every 1 in 20 posts, were identified as violent threats against New Zealand political figures. 0 personal and direct threats were detected, while 15 impersonal and direct threats were detected (35.7%), 1 personal and indirect threat was detected (2.3%), and, reflecting the hypothesis, 26 (61.9%) of the analysed threats were impersonal and indirect (fig.1).

Of all the Gab posts analysed, death threats occurred most often on posts discussing Aotearoa’s COVID-19 pandemic response- threats regarding this topic made up 80.9% of analysed threats (34 posts). Further, the *New Zealand Politics* board attracted significantly more threats (76.2% of total threats made) than the board *New Zealand Gab*, particularly on posts about ex-prime minister Jacinda Ardern. Death threats and violence-speech under ‘parent’ posts concerning Ardern across both boards made up 71.4% of total threats, whereas threats made underneath posts regarding Ardern on the *New Zealand Politics* board alone represented 87.5% of that board’s total threats (n=34). This major discrepancy in the subject of death threats and violence speech is to be expected, as surveys have shown female politicians are subject to a significantly higher amount of abuse than their male colleagues. Every-Palmer, Hansby, and Barry-Walsh found in their recent survey of MPs’ experiences of abuse that female MPs reported higher levels of abuse than their male colleagues on nearly every measure, including 46.9% experiencing threats of physical violence and 34.4% reporting receiving death threats, compared to male MPs who reported these experiences at a rate of 30% and 15% respectively .[54]

---

[54] Every-Palmer, Hansby, and Barry-Walsh, Stalking, Harassment, Gendered Abuse, and Violence, p, 5.

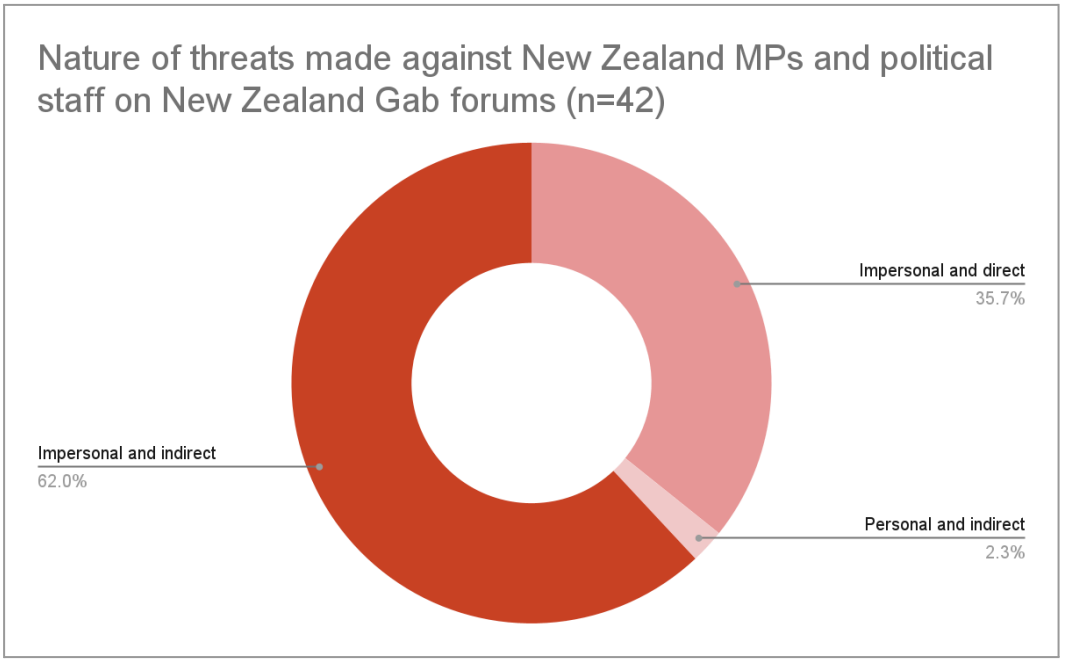

Fig. 1- The distribution of threat-types against MPs and political staff on top New Zealand GAB forums.

Paradigmatic examples are provided in order to indicate the nature of threats within each category. No personal and direct threats were recorded, so no paradigmatic examples could be provided from the source. A personal and indirect threat would have used *singular* personal pronouns such as 'I', 'me', 'my', to indicate singular, personal intent. Group pronouns such as 'we' and 'our' were taken as *impersonal,* given that the person making the threat was typically making it on behalf of some unspecified group. A personal, indirect threat found in the sample was "I will vaccinate lefties [sic] with [a handgun]". Personal intent is communicated with the personal pronoun "I" and reinforced with a definite modal verb "will". Given the unspecified target of the politically left-wing ("lefties"), the threat is categorised as being *indirect*; no specific target is indicated. In the impersonal direct category, we see examples such as "hang her!" and "she and others behind the murderous tyranny will be tried, convicted and executed for treason". In each case, the *singular* personal pronoun "she" is used to specify the MP targeted by the threat, while each poster does not personally implicate themselves in the acts they describe. In the first sample, an imperative sentence is used- while this indicates personal endorsement and personal desire to see such an act, it does not explicitly implicate the speaker in the act. In the second example, there is little connection between the poster and the specified act. While his

personal endorsement of the act is implicit rather than explicit (he says ‘hanging will happen’ rather than “hang her!”), he contributes to the fantasy of violence against politicians with a positive, celebratory tone; he describes the goals of his group to see the death of a specific politician. Finally, impersonal and indirect threats neither specify personal involvement in the violent act, *nor* indicate a specific target for the violence. This kind of threat merely expresses a desire or an approval for violence against some political enemy. Examples include: “we must kill our criminal leaders” and “they to [sic] will all be hanged for treason and crimes against humanity soon”. In the impersonal and indirect examples, we see a normative, encouraging, incitory valence in the first sample with a collective pronoun and strong modal verb “we must”, while the poster in the second sample mirrors the impersonal direct sample in merely describing what “will” happen. In each case, the speaker distances themselves from the acts described by either locating themselves as part of a group or removing themselves altogether, giving a prophetic, hopeful vision for the future. These samples also do not indicate any specified individual, rather presenting a broad class of individuals: ‘our leaders’ or ‘they’, which in the ‘parent’ post is specified as the ‘ruling class’. Now that the nature of each category and the violent communications therein are specified, the implications of this study may be understood more thoroughly.

The hypothesis, we see, is strongly supported. The results indicated by this short study align with results found in other political environments, such as the United States,[55] indicating that indirect and impersonal death and violence-threats have a *major* role to play in communities of hate, creating an environment where violence is normalised, accepted, and endorsed as part of the social lexicon. [56,57] Further, the saturation of impersonal death threats in this community demonstrates that the majority of this problematic speech is *incitory* in nature. Impersonal threats do not communicate personal intent, instead functioning to encourage others to commit acts of violence. In what follows, the tools parliament has at its disposal for addressing the harms caused by this kind of public speech- particularly as it pertains to the incitement of violence- are considered, specifically options for legislative reform.

Legislative Solutions Compared Across the Tasman

[55] Eisler, Szep, So, and Hart, ‘Anatomy of a Death Threat’.
[56] Bailard, Tromble, Zhong, Bianchi, Hosseini, and Broniatowski, “‘Keep Your Heads Held High Boys!,” pp. 2, 14.
[57] Berentson-Shaw and Elliott, ‘Online Hate and Offline Harm,’ pp. 14-15.

Individuals subject to these threats have emphasised that the current legislative landscape is unequipped to deal with the kind of death threats we most frequently see in online communities- that is, *incitory* threats. Remembering the sentiment expressed by the agencies unable to address the 'Nuremberg 2.0' website, here we investigate Aotearoa's online communication legislation, comparing it to equivalent legislation in Australia, to understand the potential and precedent for improvement.

New Zealand's Harmful Digital Communications Act 2015 (HDCA) is intended to "deter, prevent, and mitigate harm caused to individuals by digital communications; and provide victims of harmful digital communications with a quick and efficient means of redress".[58] Towards this end, the HDCA sets out 10 *Communication Principles*, breaches of which by internet users can be investigated and addressed by courts.[59] While the principles prohibit the incitement to *online harassment* and incitement to *suicide*, no such principle exists to prohibit the incitement to *offline* violence.[60] Further, individuals may not bring any person to court under the HDCA unless they have attempted mediation,[61] a process which one academic subject to public death threats has described as dangerous, since the process satisfies the harassers, showing that "they are having an effect".[62] Under the HDCA, then, it is an offence to make a communication with the intention of causing harm against an individual,[63] but it is *not* an offence to make a communication with the intention (or effect) of bringing about *offline harm* against an individual. Criticism has been brought against the HDCA that is not fit for purpose in a landscape where death threats are made in public forums, rather than directly to individuals. As the same academic notes, the Act "seems to have been written for teenagers experiencing cyberbullying".[64] The 2020 Royal Commission Inquiry into the Christchurch mosque attacks similarly criticised the Act's requirement for offending speech to specify a *victim*, noting that charges could not be brought against individuals who "denigrate groups rather than particular individuals".[65] Results of this Gab study reflects this- most instances of incitement to violence (i.e. impersonal threats) were indirect; against unspecified groups like 'the elites' and 'our leaders'. However, death threats and incitement against the groups of interest to the Christchurch Royal Commission report

---

[58] *Harmful Digital Communications Act* 2015 (NZ) pt 1 sub-pt 1 s 3.
[59] *Harmful Digital Communications Act* 2015 (NZ) pt 1 sub-pt 1 s 6 (1-2), pt 1 sub-pt 2 s 12 sub-s 12.2(a).
[60] *Harmful Digital Communications Act* 2015 (NZ) pt 1 sub-pt 1 s 6(1).
[61] *Harmful Digital Communications Act* 2015 (NZ) pt 1 sub-pt 2 s 12 (1), pt 1 sub-pt 2 s 13 sub-s2(a).
[62] Roberts, 'What it Took to Stop Harassment'.
[63] Harmful Digital Communications Act 2015 (NZ) pt 1 sub-pt 2 s 22 sub-pt 1(a).
[64] Roberts, 'What it Took to Stop Harassment'.
[65] Royal Commission of Inquiry into the Terrorist Attack On Christchurch Mosques on 15 March 2019, 'Part 9: Embracing Social Cohesion and Diversity'. *Royal Commission of Inquiry into the Terrorist Attack On Christchurch Mosques on 15 March 2019* 3 (1) 2020, p. 712.

(groups with a 'protected characteristic') were not measured in this study. Equivalent pieces of legislation in our pacific community might point to options for reform for Aotearoa, to protect both our MPs and our other vulnerable communities from the harm from public online death threats.

Australia's *Basic Online Safety Expectations* (BOSE) set out in their Online Safety Act 2021 serve a similar function to Aotearoa's *Communication Principles* set out under the HDCA. The BOSE, though, more thoroughly and specifically address online violence and abuse. For example, social media service providers must minimise the amount of content that "promotes", "incites", or "instructs in abhorrent violent conduct".[66] They also must have clear and accessible means by which users can report content that "promotes", "incites", or "instructs in abhorrent violent conduct".[67] Aotearoa's Communication Principles do not contain prohibitions of this type. While they recognise the harm that can be done by publicly disclosing sensitive personal facts or spreading false allegations, they do not recognise the harm done to an individual through direct incitory threats of the kind so prevalent on New Zealand Gab forums.

A major difference between the Online Safety Act 2021 and the HDCA, though, is that the former only addresses what *platforms* can and can not host, while the HDCA legislates against what *individuals* can post online. Laws regulating online platforms are valuable, but they can not alone address the problem of online incitement turning into offline violence, or fear thereof. When people get censored on one platform, they simply move to another or create their own, creating a cat-and-mouse game of hateful platforms. Indeed, Gab was established as an alternative to X (formerly Twitter) for this very purpose.[68] A strength of the HDCA, then, is that it allows us to hold *individuals* accountable for the violent and threatening communications they themselves create. We see both a model and a precedent for improving personal accountability for violent online communications (rather than platform accountability) in legislation regarding violent communications to *groups*. Australia's Criminal Code Act 1995 prohibits individuals from urging (and intending their urging to cause) "a group" to commit violence "against a group" with a protected characteristic.[69] A similar law exists prohibiting someone from urging a *group* to commit violence against an individual by virtue of that individual's protected characteristic.[70] These provisions remain imperfect, though. First, *being a political official* is not a protected class, (and indeed it ought not be so

---

[66] *Online Safety Act* 2021 (Cth) pt 4 sec 45 div 2 s 46 sub-s 1(c) v, vi, vii.
[67] *Online Safety Act* 2021 (Cth) pt 4 sec 45 div 2 s 46 sub-s 1(e) vi, vii, viii.
[68] St.Aubin and Stocking, 'Key Facts About Gab'.
[69] *Criminal Code Act* 1995 (Cth) ch 5 p 5.1 div 80 sub-div C 80.2A sub-s 1(a-c).
[70] *Criminal Code Act* 1995 (Cth) ch 5 p 5.1 div 80 sub-div C 80.2B sub-s 1(a-d).

as to protect genuine criticism and dissent), so these laws would not necessarily prevent communications that incite violence against MPs or political staff as a group. Second, these sections of the legislation do not specify their application to online communications. Nonetheless, these prohibitions in tandem with the BOSE create the possibility for agencies to take action against sites saturated with impersonal indirect, or impersonal direct threats of violence and murder against politicians with a *specific ideological adherence,* given the inclusion of "political opinion" as a protected characteristic.[71]

Aotearoa's equivalent legislation, the Crimes Act 1961, does not seem as equipped to deal with situations where a group or the public is incited to commit an act of violence against a group or an individual. The Act renders someone *party* to, (and therefore guilty of [72]), an offence they incited if the offender was likely to or actually did commit the offence.[73] The likelihood of the offence occurring likely rules out all online group communications, given that the inciter has no means for determining how likely or capable their audience is to commit that offence. The Crimes Act also contains a provision against inciting murder, but it pertains to individual-on-individual communication rather than impersonal acts of incitement to a group that are, as we see, so common on fringe social media platforms.[74]

We see that the Crimes Act and the HDCA are ill-equipped to handle the phenomenon of public 'hit lists' like 'Nuremberg 2.0' and public online incitement to violence against political officials. Including provisions against the 'promotion', 'incitement' or 'instruction' to commit violent acts in the Communication Principles of the HDCA may open up avenues for individuals to seek justice when there has been a "serious" or "repeated breach" of the Communication Principles.[75] There is also a significant and precedented opportunity, when we compare our Crimes Act to the Criminal Code Act, to include provisions against group or public 'promotion', 'incitement' or 'instruction' to commit violent acts against individuals or groups. However, using the HDCA's infrastructure to achieve this goal could result in accountability *both* for social media platforms *and* for the individuals who make "serious" and "repeated" incitement and death threats hosted thereupon.[76]

Closing Remarks

---

[71] *Criminal Code Act* 1995 (Cth) ch 5 p 5.1 div 80 sub-div C 80.2A sub-s 1(c), 80.2b sub-s 1(d).
[72] *Crimes Act* 1961 (NZ) pt 4 sub-pt 66 sub-s 1(d).
[73] *Crimes Act* 1961 (NZ) pt 4 sub-pt 70(1).
[74] *Crimes Act* 1961 (NZ) pt 8 sub-pt 174.
[75] *Harmful Digital Communications Act* 2015 (NZ) pt 1 sub-pt 2 s 12 sub-s2(a).
[76] *Harmful Digital Communications Act* 2015 (NZ) pt 1 sub-pt 2 s 12 sub-s2(a).

Hate and violent threats against Members of Parliament and their staff is on the rise, and social media has a significant role to play in this phenomenon. It is not just the direct threats we ought to be concerned about, though; we must pay attention to the online communities that forment and incite this hatred. Taking a focus on one of New Zealand's largest, publicly accessible extremist social platform Gab, it was shown that death threats against politicians make up over 5% of posts, with the majority being incitory in nature. If this is the case across other online extremist communities, there is a clear need for legislative solutions. One legislative solution explored in this paper was an addition to the *communication principles* set out in the Harmful Digital Communications Act 2015 that prohibits communications inciting a group to violence, but more possibilities are open to us, such as an extension of the definition of 'objectionable' content in the Films, Videos, and Publications Classification Act 1993, or a widening of the 'racial disharmony' clause in the Human Rights Act 1993 to include all protected groups.

This study was small, and there were a number of limitations. Primarily, replacing or supplementing the single junior researcher with a more experienced researcher would yield broader, higher-quality results. The small number of posts analysed (n=810) may not have produced a representative sample, and the qualitative nature of single-researcher coding may have produced biassed results. Solutions for this issue involve multi-researcher coding and establishing a broader scope for data gathering, particularly the inclusion of more fringe social media platforms such as Telegram.

Future research into the nature of violent fringe social media communities that produce and foment violence against political figures ought to be conducted with a wider scope, and on a wider array of platforms, especially Facebook and X (formerly Twitter) where a majority of New Zealand's extremists congregate.[77] Eisler et. al's data gathering method of collecting death threats made directly to the offices of politicians and political staff could be repeated here, too, and coded based on their personal/impersonal nature to reveal the amount of personal responsibility for violence people are willing to take when not in the comfort and anonymity of their extremist community. Linguistic, political, and cultural variations on the personal/impersonal direct/indirect death threat trend could also be investigated to uncover whether cultural differences change the nature of death threats. If they do, it might reveal extra-legislative possibilities for addressing the issue of incitement to violence via cultural

[77] Comerford, Guhl, and Miller, 'Understanding the New Zealand Online Extremist Ecosystem'.

initiatives, like expansions of the Department of Prime Minister and Cabinets' *Preventing and Countering Violent Extremism Fund*.[78]

[78] Department of Prime Minister and Cabinet. 'Preventing and Countering Violent Extremism Fund'. Accessed at: https://www.dpmc.govt.nz/our-programmes/national-security/counter-terrorism/preventing-and-countering-violent-extremism-fund.