

# Measuring trends in online hate speech victimisation and exposure, and attitudes in New Zealand

Prepared by Dr. Edgar Pacheco & Neil Melhuish

## What is this about?

On 15 March 2019 a devastating terrorist attack against the Al-Noor Mosque and the Linwood Islamic Centre in Christchurch killed 51 people and injured 49, some seriously. In addition to the tragic loss of life, the attack generated one of the most significant online safety challenges New Zealand has experienced. In the immediate aftermath a text reportedly written by the attacker that included racist and discriminatory language and a livestream video of the attack went viral and continued to reverberate across the internet.

About 46 minutes after the attack began, Netsafe received its first call from the public about this content being accessible online. It went on to receive just under 600 enquiries and complaints in the days following. While most of these were about content showing the shootings, many were from people complaining they had received or witnessed hateful digital communications inspired by the attack. The Netsafe contact centre team worked closely with its trusted government and industry partners to help remove video and other related content from the internet.

In response to the way that social media was used as a tool to distribute images of the attacks, the New Zealand and French governments joined to bring together Heads of

## Summary of findings

- Overall, 15% of New Zealand adults reported having been personally targeted with online hate speech in the last 12 months.
- Compared to our 2018 survey, this result is higher by 4 percentage points.
- Over one third of personal experiences of online hate speech occurred after the Christchurch attacks.
- Half of Muslim respondents said they were personally targeted with online hate in the last 12 months. Prevalence was also more common among Hindus.
- Similar to 2018, people with disabilities and identifying as non-heterosexuals were also targeted at higher rates.
- About 3 in 10 adult New Zealanders say they have seen or encountered online hate speech content that targeted someone else.
- Nearly 7 in 10 New Zealand adults think that online hate speech is spreading.
- Over 8 in 10 adults believe that social media platforms should do more to stop online hate speech.
- While three-quarters would support new legislation to stop online hate, a similar proportion considers that more than that is needed to prevent its spread.
- At the same time, a large majority, 8 in 10, believe that everyone has a role to play in addressing hateful speech.
- More than half disagreed with the idea that people should be entitled to say whatever they want online. A quarter do not have an opinion.

State and Government and leaders from the technology sector to adopt the Christchurch Call<sup>1</sup>. The Christchurch Call is a commitment by governments and technology companies to eliminate terrorist and violent extremist content from the internet. Netsafe is a member of the Christchurch Call Advisory Network of civil society organisations, reflecting its engagement in tackling related issues and its commitment to continuing to share its experience and expertise.

In 2018 we conducted a study to explore adult New Zealanders' personal experiences of, and exposure to, online hate speech. The results from that research<sup>2</sup> provided the first nationally representative measure of the extent and nature of online hate speech in New Zealand.

Following the Christchurch attacks, we decided to build on and extend last year's study. As a result, we have prepared this report which presents the findings from our 2019 study. The report provides valuable insights to inform the current public debate and academic analysis on the topic of online hate speech in the aftermath of the Christchurch attacks. Our main objective was to examine trends in the online hate speech experiences of New Zealanders by comparing results from 2018 and 2019. To the best of our knowledge, this is the only study that compares annual trends in online hate in New Zealand.

A contribution of this study, compared to last year's research, is that we took a closer look at religion as a relevant demographic variable to measure online hate. Another innovation is that this study explored public perceptions about issues related to online hate speech, such as whether they think hateful online content is growing in New Zealand and around the world.

Netsafe carries out research as a wider part of its statutory role as approved agency under the Harmful Digital Communications Act 2015 to inform the design and delivery of its resources

and services and provide research-based evidence for others working to address online safety issues. The topic of this report sits within the scope of the Act's ten communications principles that together describe a range of potentially harmful types of online communications<sup>3</sup>. Specifically, principle 10 extends to a limited set of communications which could be considered hateful in nature<sup>4</sup>.

Netsafe's extensive and original research programme also includes collaborations with international partners. As part of this study we shared background information and our questionnaire with the Australian Office of the eSafety Commissioner to enable comparison between the experiences of the two countries' internet users. The results of this will be published in early 2020.

It is our hope that the insights in this report contribute relevant evidence to inform discussion about sensitive issues related to online hate speech in New Zealand. We also believe that this work usefully adds to the growing international body of research on this topic. Finally, the delivery of this report reflects Netsafe's broad commitment to government, industry, academics and others to address online safety issues with research-based evidence about online behaviours and the impact of digital communications in New Zealand.

## What we know so far

Despite the importance of understanding the extent of online hate speech, there is little published data on the topic in New Zealand. Among the few available studies, research conducted by Netsafe in 2017 regarding harmful digital communications found that in the prior year 9% of participants received a digital communication that said offensive things about their lifestyle, or their religious or political beliefs (Pacheco & Melhuish, 2018a). Another study on racial abuse online (see ActionStation, 2019)

<sup>1</sup> <https://www.christchurchcall.com>

<sup>2</sup> Online Hate Speech: A Survey on Personal Experiences and Exposure Among Adult New Zealanders <https://ssrn.com/abstract=3272148>

<sup>3</sup> See Harmful Digital Communications Act 2015, s 6(1).

<sup>4</sup> Principle 10: A digital communication should not denigrate a person's colour, race, ethnic or national origins, religion, gender, sexual orientation or disability.

focused on its prevalence among Māori, Pacific and Asian people. However, both studies, while useful, provide a narrow definition of hate speech and/or limited evidence about the extent of this phenomenon.

The first nationally representative study on the topic was conducted a year later by Netsafe. The research sought to measure adult New Zealanders' personal experiences of and exposure to hateful speech in the prior year. Overall, Netsafe's study found that 11% of respondents were personally targeted with online hate with higher rates among minority ethnic groups, younger adults, males, people with disabilities, and non-heterosexual respondents. Similarly, the study found that religion followed by political views and appearance were the most common perceived reasons for hate speech victimisation (see Pacheco & Melhuish, 2018b). In regard to exposure, about 3 in 10 of adult New Zealanders indicated having seen hateful speech that targeted someone else (Pacheco & Melhuish, 2018b). Although the study did not collect data based on people's religious affiliation, it provides robust measures and personal accounts of emotional impact, concluding that online hate speech in New Zealand is more likely to target minorities (Pacheco & Melhuish, 2018b).

On the other hand, international survey-based research on online hate speech has centred on sampling teenagers and young adults in the context of exposure to hateful content (Pacheco & Melhuish, 2018b). In the same line of inquiry, a recent Netsafe study on the experiences of online risks and perceived harm among New Zealand children aged 13-17 provides useful evidence on the matter. The study reveals that 27% of teenagers had seen websites or online discussions where people talk about hate messages attacking certain groups (e.g. people of different colours, religions or nationalities). What is more, online hate exposure was more prevalent among teenage girls and teens aged 17 years old (see Pacheco & Melhuish, 2019). As

this finding is part of a larger research project<sup>5</sup> regarding different types of online risks, more research is still required to understand in depth New Zealand children's experiences of and the impact of hateful speech.

All the insights highlighted above broadly depict the extent of online hate speech in New Zealand. However, as government agencies are not currently required to systematically collect data on this type of incident (Greenfield & Menon, 2019; Spoonley, 2018), there is a lack of longitudinal evidence of trends in this respect. Also, apart from measuring the prevalence of personal experiences of and exposure to online hate, another knowledge gap regards our limited understanding of public attitudes towards different aspects related to hateful online content. With only a handful of New Zealand-based empirical investigations, there is, thus, a need for more research and robust evidence; otherwise, intended policies and responses to prevent hate speech might be at risk of falling short. The aspects highlighted here are critical especially in the aftermath of the Christchurch attack. We are confident the insights provided in this study will help to inform public conversation and the implementation of strategies and activities to ensure all New Zealanders, no matter what their characteristics, are safe online to benefit from the multiple opportunities of the digital environment.

## What we did

Data used for this report came from Netsafe's 2019 Annual Population Survey (APS)<sup>6</sup>. The research questions that guided the study were:

- What are the personal experiences of adult New Zealanders regarding online hate speech in the prior year? How do these experiences compare with our 2018 results?
- What is the extent of exposure to online hate speech among adult New Zealanders? How do findings about exposure compare with our 2018 results?

<sup>5</sup> See <http://globalkidsonline.net/new-zealand/>

<sup>6</sup> The APS is a large quantitative study that explores the interaction between adult New Zealanders and digital technologies in the context of the Harmful Digital Communications Act 2015. See <https://www.netsafe.org.nz/annual-population-survey-2017/>

- What are the attitudes of adult New Zealanders regarding issues related to online hate speech?

To measure personal experiences of and exposure to online hate speech, we reapplied questions from the survey tool used in the 2018 study<sup>7</sup>. In addition, for this year, we included a question that asked those people that told us they had received a hateful communication in the last 12 months when this incident occurred. We also added a question asking participants about statements regarding online hate-related issues. We did so to measure people's views about the role and/or approaches of government, social media platforms, and people themselves to deal with this sort of online content. The statements added in the survey reflected public discussion and interest at the time of developing the research instrument and following the Christchurch attacks.

Our working definition of online hate speech followed the conceptualisation developed for last year's study on the basis of international literature. Hence it is defined as:

*“any technology-mediated speech or digital communication that offends, discriminates, denigrates, abuses and/or disparages a person(s) on the basis of a group-defining characteristic such as race, ethnicity, gender, nationality, sexual orientation, religion, age, disability, and others”*

A total of 1,161 New Zealanders aged 18 and older completed an online survey. Data collection was conducted between 4 and 26 June 2019 by Colmar Brunton (data for last year's study was collected during June 2018). The research sample is representative of the population in terms of age, gender, ethnicity, and religion<sup>8</sup>. Regarding gender, females represented 51.7% of the total sample, males made up 47.9%, and 0.4% of participants identified as gender diverse. In terms of ethnicity our sample was distributed as follows: NZ European/Pākehā (71%), Māori (12%), Pacific

(5%), Asian (12%), and other ethnicities (10%). In terms of age, the sample was distributed as follows: 21% were participants aged 18-29 years old, while those aged 30-39, 40-49, and 50-59 represented 16%, 19%, and 17% respectively. Finally, 14% were 60-69 and 13% were 70 years or older.

Colmar Brunton conducted booster interviews with people who described their religion as Buddhist, Hindu, or Islam/Muslim to ensure we had enough samples of these groups for analysis. The results were weighted by the incidence of each religion in the population so that the booster interviews did not affect the representativeness of the overall results.

Data about sexual orientation was also collected: 6% of participants identified themselves as gay, lesbian, bisexual, or other. These respondents were grouped as non-heterosexual.

The overall margin of error for the 2019 results is +/- 3.1% at 95% confidence. Meanwhile, +/-4.2 is the maximum margin of error associated with comparing survey results between 2018 and 2019 at the 95% confidence level.

Note that percentages in figures and tables may not total exactly 100% due to rounding or because survey participants were allowed to choose multiple answers to some questions.

Due to the sensitive nature of the topic and the timing of the fieldwork, we put special care into managing ethics. For example, we did not include any reference to the Christchurch attacks to avoid unnecessary triggering of memories and emotions associated with the events. Similar to last year's study, we provided all participants with information about the nature and goals of the study in advance. We also asked them for their consent to take part in the study and guaranteed that their responses were confidential and that their data will be kept protected. Similarly, Colmar Brunton followed the Research Association's Code of Practice

<sup>7</sup> For details about the questions asked and methodology see our 2018 report *Online Hate Speech: A survey on personal experiences and exposure among adult New Zealanders*. <https://www.netsafe.org.nz/online-hate-speech-report/>

<sup>8</sup> Our methodology enabled us to collect data from a representative sample of New Zealand's four largest religious groups: Christian/ Māori Christian, Buddhist, Hindu, and Islam/Muslim.

and ensured that participants had the right to withdraw at any time during the study. On-screen links to relevant services, including Netsafe’s help service, were also provided to participants when they were responding to the survey. We also included an email address so those participants who required further information about the project were able to reach Netsafe’s Policy & Research team.

Social research has its challenges and limitations, and this has been the case for the present study. First, despite expanding last year’s online hate speech questionnaire (e.g. by adding a question about attitudes regarding online hate issues), the study still provides a useful but broad picture of the topic in New Zealand. While including an open-ended question for this year’s study was a suitable response to the challenge, we encourage researchers to explore online hate in more depth on the basis of qualitative inquiry as well as computational methods.

Another challenge is that this survey-based research collected data from a specific point in time (June 2019). Thus, we acknowledge that public opinion about sensitive issues such as online hate might change over time.

This study describes findings regarding people with disabilities and non-heterosexual people. A note of caution is that results about these two groups should be considered as indicative rather than representative. Netsafe believes that it is critical to collect evidence regarding hard-to-reach population groups. We encourage other researchers to consider the needs and experiences of vulnerable groups when researching experiences of online abuse and harassment.

We also acknowledge that defining online hate speech is challenging. It is not only an evolving concept (Pacheco & Melhuish, 2018b) but the lack of methodological consensus on how to measure it limits its full understanding and the responses to address it promptly and effectively. We are confident that our approach to measuring online hate speech, which captures prior research, will guide future research on the topic.

## What we found

This section presents the main findings of our 2019 Online Hate Speech survey module. Note that some of the figures include findings from our 2018 study in order to identify trends in personal experiences and exposure to hateful speech.

---

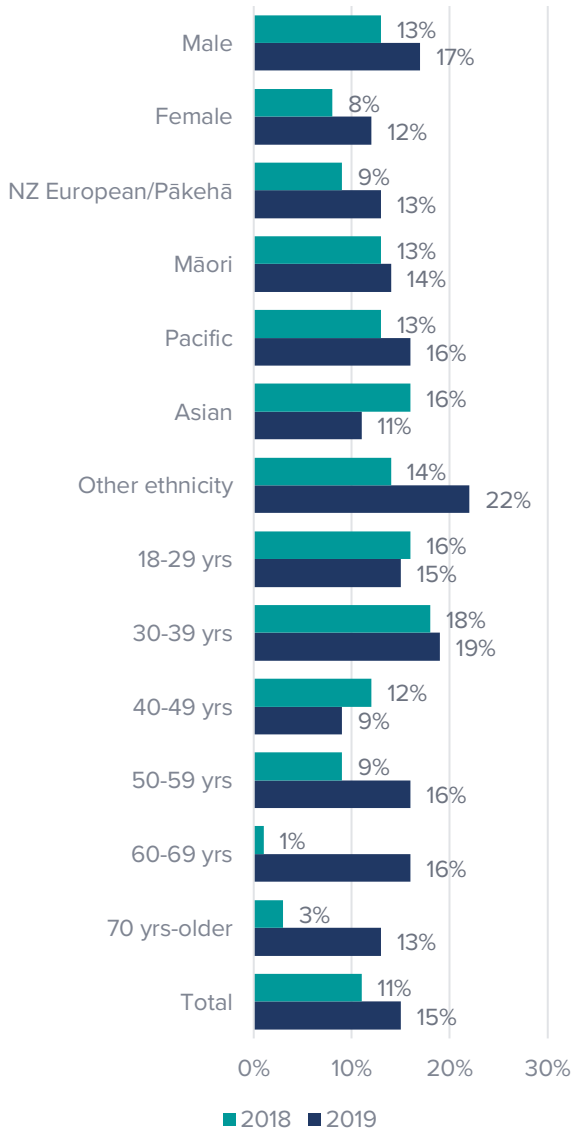
### BEING THE TARGET OF ONLINE HATE SPEECH

We asked participants whether they have personally received a digital communication that offended, discriminated against, denigrated, abused and/or disparaged them because of their race, ethnicity, gender, nationality, sexual orientation, religion, age, disability, and/or other characteristics. Our results show that 15% of respondents said they were targeted with hateful speech in the prior year. Although a minority was personally targeted, the prevalence of online hate speech in 2019 is higher by 4 percentage points compared to our 2018 survey (11%) – see details in Figure 1.

As Figure 1 also illustrates, similar to last year’s study, males were more likely to report personal experience of online hate compared to females.

However, in terms of ethnicity there were some interesting results. For example, the prevalence of hate speech among participants who identified themselves as an “other ethnicity” increased from 14% in 2018 to 22% in 2019. In contrast, the incidence of hateful online content among Asian participants decreased from 16% in 2018 to 11% in 2019. Percentages among younger adults remained unchanged overall.

Some differences were also found regarding age. Rates of online hate speech among older adults and seniors (50 years old and over) increased significantly in 2019 compared to 2018. For example, in 2018 only 1% of 60-69-year-old participants reported having been the target of hateful speech. In 2019, the incidence of online hate speech for this group raised to 16%. See Figure 1.



**Figure 1. Prevalence of personal experiences of online hate speech in 2018 and 2019**

Q40 - In the last 12 months, have you received a digital communication that offended, discriminated, denigrated, abused and/or disparaged you because of your personal identity/beliefs? (e.g. race, ethnicity, gender, nationality, sexual orientation, religion, age, disability, and/or other)

Base: All participants aged 18 and over (2018 n=1,001, 2019 n=1,161)

Regarding religion, 52% of participants who identified as Muslim indicated being the target of online hate in the past 12 months. This was also the case for 32% of Hindus. Meanwhile, 13% of participants whose religious affiliation is Christianity reported having received hateful speech. A very small percentage (3%) was reported by Buddhist participants.

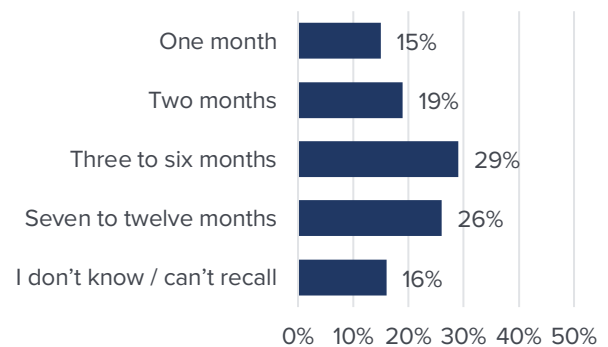
Another interesting finding relates to disability. The rate of online hate speech among people with disabilities rose by 10 percentage points

compared to last year's survey (from 15% to 25%).

Regarding participants' sexual orientation, about 15% of those who identified themselves as heterosexual and 23% of non-heterosexual participants indicated having received hateful digital communications one or more times in the last 12 months. In contrast, the rates in 2018 for heterosexual and non-heterosexual respondents were 9% and 26%, respectively.

## OCCURRENCE OF ONLINE HATE SPEECH

Participants who reported being personally targeted with online hate (n=171) were asked a follow-up question regarding when the incident happened. Specifically, they answered if the experience occurred in the last "month", "two months", "three to six months", or "seven to 12 months". They also could answer "I don't know/can't recall". As previously indicated, the survey was conducted between 4 and 26 June 2019. Our results show that 34% of participants said they were targeted with hateful content in the previous two months, 29% in the last three to six months and 26% in the seven to twelve months prior to the survey. See Figure 2.



**Figure 2. When the personal experience of online hate speech happened in 2019**

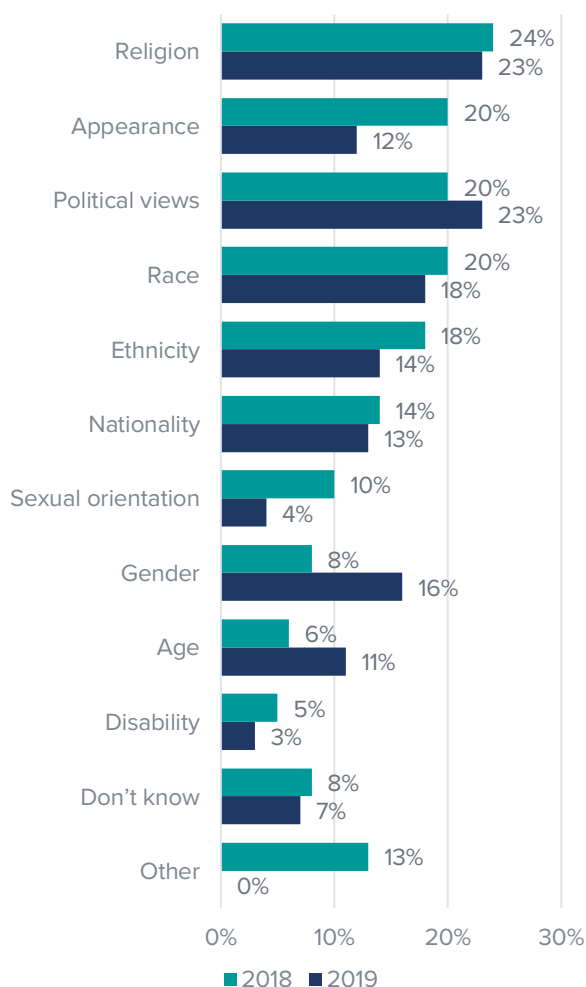
Q40A – More exactly, can you tell us if you received this type of digital communication in the last... Please choose all that apply.

Base: Participants who were personally targeted with hateful content in 2019 n=171

Note: New question in 2019

## PERCEIVED REASONS FOR BEING THE TARGET OF ONLINE HATE SPEECH

As in 2018, this year we asked participants about their perceptions of why they think they were personally targeted with hateful speech. Our findings show that the most common reasons were religion, political views, race, and gender – see Figure 3. Like last year’s survey, religion (23%) remained the main perceived reason for being targeted with online hate, with political views equally rated in importance (23%) this year. Some noteworthy changes include gender being given as a reason, as it doubled from 8% in 2018 to 16% in 2019. Also, in 2018 the reason ‘appearance’ was among the top perceived reasons for online hate, but it went down from 20% to 12% in 2019.



**Figure 3. Perceived reasons for receiving online hate speech in 2018 and 2019**

Q41 - The digital communication(s) I received targeted me because of my...  
Please choose all that apply in relation to your online experiences in the last 12 months.

Base: Participants who were personally targeted with hateful content (2018 n=112, 2019 n=171)

## EXPOSURE TO ONLINE HATE SPEECH

To measure exposure to online hate speech, we asked all participants whether in the last 12 months they had seen or been exposed to a digital communication targeting someone else or a group because of their race, ethnicity, gender, nationality, sexual orientation, religion, age, disability, and/or other. As Figure 4 illustrates, overall, 28% of participants (nearly 3 in 10) reported exposure to hateful content one or more times. This percentage remains unchanged in this year’s survey.

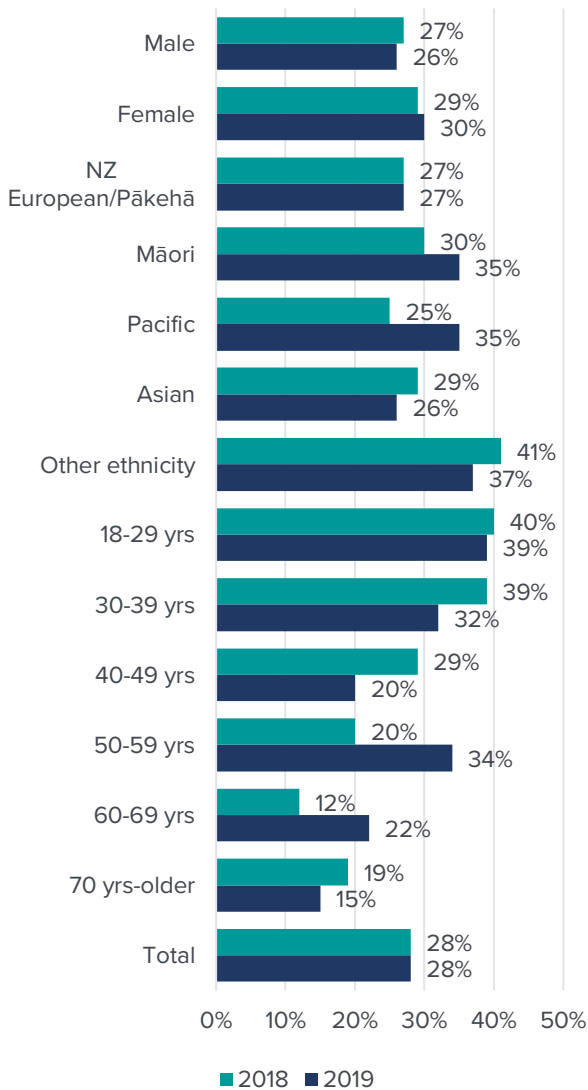
Also, compared to 2018, our 2019 results show that exposure to online hate among Pacific and Māori participants increased by 10 and 5 percentage points, respectively. In contrast, the incidence of hateful speech exposure slightly decreased this year for Asians and those in the “other ethnicity” group. However, as in 2018 people in the latter group reported the highest rate of exposure compared to other ethnic groups.

On the other hand, our data depict an increase of exposure for some specific age groups. This is the case for 50-59-year-old participants whose reported exposure jumped to 34% from 20% in 2018. Similarly, 22% of participants aged 60-69 indicated being exposed to hateful speech which is higher by 10 percentage points in relation to the 2018 study (12%). While results were much the same for 18-29-year-olds in both years, there was a decrease for adults aged 30-49. In this respect, rates of exposure among those aged 30-39 (32%) and 40-49 (20%) were lower compared to 2018: 39% and 29%, respectively.

In terms of gender, females reported a higher incidence of online hate exposure, as in 2018. However, rates were slightly higher in this year’s results compared to those from last year.

When looking at the results in the context of religious affiliation, the highest rate of exposure to online hate speech was reported by Muslim participants (67%) followed by Hindus (37%). Meanwhile, rates of exposure for Christian and Buddhist participants were 25% and 20%, respectively.

For people with disabilities (34%) exposure to online hate was more common than for non-disabled participants (27%).



**Figure 4 Exposure to online hate speech in 2018 and 2019**

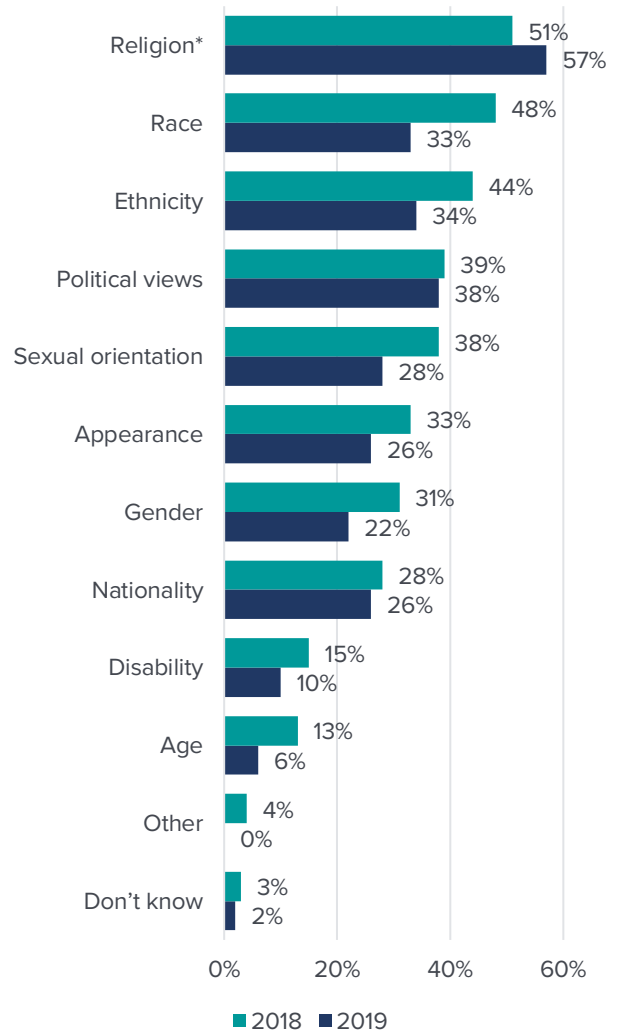
Q43 - In the last 12 months, have you seen or been exposed to a digital communication that targeted someone or a group because of their race, ethnicity, gender, nationality, sexual orientation, religion, age, disability, and/or other?

Base: All participants aged 18 and over (2018 n=1,001, 2019 n=1,161)

## PERCEIVED REASONS FOR ONLINE HATE SPEECH AGAINST OTHERS

Participants who said they were exposed to online hate (n=339) were asked what they thought were the reason(s) others were targeted. As Figure 5 shows, and like 2018, religion, political views, ethnicity, and race were the most commonly perceived reasons others were the target of hateful online speech. In terms of ethnicity, all groups indicated that religion was the main reason for exposure but

Pacific, Asian and Māori rated race and ethnicity higher. Interestingly, females' responses (30%) in relation to citing gender as a reason for online hate exposure were much higher than for males (11%). For participants who identified themselves as non-heterosexual, sexual orientation (66%) was the most common reason for exposure to online hate.



**Figure 5. Perceived reasons for online hate speech against others in 2018 and 2019**

Q44 - The digital communication(s) I have seen or been exposed to targeted other(s) because of their... Please choose all that apply in relation to what you have seen online in the last 12 months.

Base: Participants aged 18 and over who indicated to have been exposed to hateful content (2018 n=285, 2019 n=339)

\* Caution: Sample size for Muslim and Hindu groups <30 each

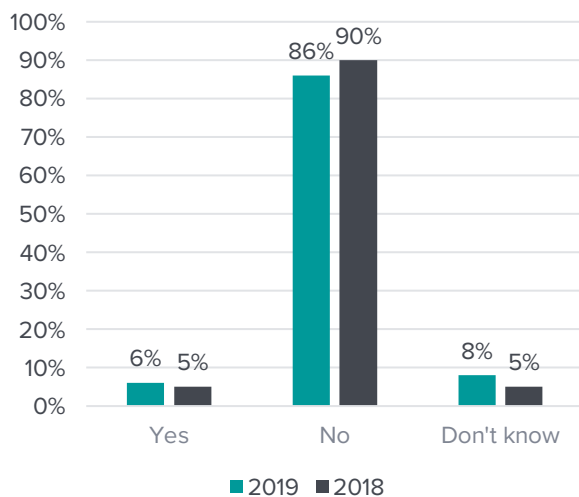


## ENGAGEMENT WITH ONLINE SITES PROMOTING/DISTRIBUTING ONLINE HATE SPEECH

As in 2018, we asked participants whether in the prior year they had intentionally visited a website, online forum and/or social media group that targets people because of their race, ethnicity, gender, nationality, sexual orientation, religion, age, disability, and/or other characteristic. Overall, a very small number of participants (6%) intentionally engaged with this type of online environment — see Figure 6.

In terms of demographics, engagement with sites promoting online hate speech was twice as common among males (8%) than females (4%). It was also more common among those who identified as an “other ethnicity” (13%), Pacific people (12%), and Muslims (25%).

As described in last year’s report, our questions about intentionally visiting an online site(s) that promotes or distributes hateful speech do not explain, for instance, motivations for visiting it or alignment with the ideology promoted. Thus, further interpretation should be considered cautiously.



**Figure 6. Overall engagement with online sites promoting online hate speech in 2018 and 2019**

Q45 - In the last 12 months, have you intentionally visited a website, online forum and/or social media group that targets people because of their race, ethnicity, gender, nationality, sexual orientation, religion, age, disability, and/or other?

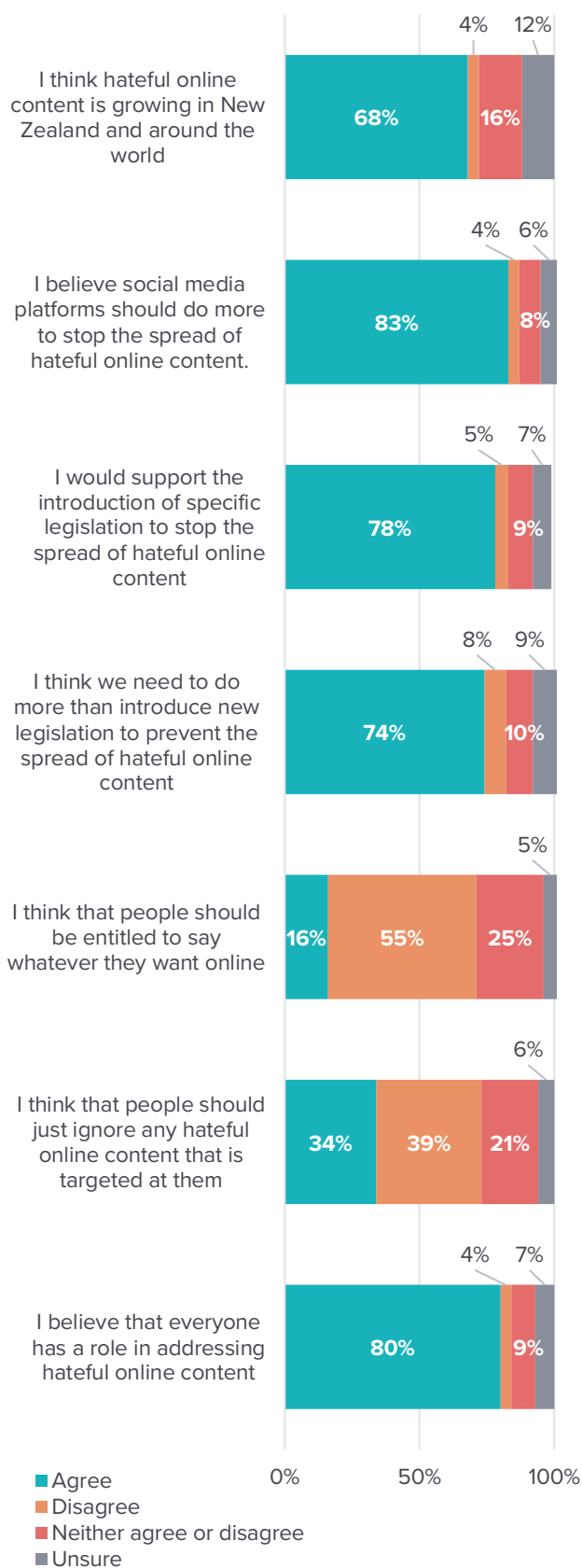
Base: All participants aged 18 and over (2018 n=1,001, 2019 n=1,161)

## LEVEL OF AGREEMENT WITH ONLINE HATE-RELATED ISSUES

For this year’s survey we added a new question that sought to explore survey participants’ level of agreement with seven statements related to online hate speech issues. The following Likert scale was used to measure their responses: “strongly disagree”, “disagree”, “neither agree nor disagree”, “agree”, “strongly agree”, and “unsure”. Figure 7 presents aggregated findings.

The first statement listed in the question was “I think hateful online content is growing in New Zealand and around the world”. As Figure 7 shows, nearly 7 in 10 participants (68%) agreed with it and a very small percentage disagreed (4%). While rates did not vary substantially in terms of gender and age, there were some interesting results related to ethnicity. In this respect, the highest level of agreement with the statement was reported by Pacific participants (79%) followed by Māori respondents (75%). Participants experiencing a disability (75%) were also more likely to agree with the statement than non-disabled respondents (69%). On the other hand, those under the age of 30 (14%), and participants whose religious affiliation is Hindu (19%) or Muslim (12%) were more likely than average to disagree.

The second statement was related to the role of social media platforms. A large majority (83%) indicated that social media companies should do more to stop the spread of online hate speech. Only 4% disagreed. Participants aged 70 and over (92%) as well as 40-49-year-olds (88%) tended to agree more with the statement. Meanwhile, participants who identified themselves as Muslim (13%) were more likely than average to disagree with the statement.



**Figure 7. Level of agreement with statements related to online hate speech issues**

Q46 - Now, how much do you agree or disagree with the following statements...

Base: All respondents in 2019 survey (1,161)

Note: new question in 2019

When asked if they would support the introduction of specific legislation to stop the spread of hateful online content, nearly 8 in 10 participants (78%) agreed. Females (85%) were significantly more likely to agree with the statement than males (71%). Similarly, in terms of religious affiliation, Hindus (88%) followed by Muslims (84%) were more likely to agree. In contrast, non-heterosexual respondents (10%) were more likely than average to disagree with the statement.

Meanwhile, three quarters of all respondents (74%) agreed that there is a need to do more than introducing new legislation to prevent the spread of online hate. This belief was higher among those aged 70 and older (87%) followed by 30-39-year-olds (79%). The lowest level of agreement was reported by young participants aged 18-29 (68%). Also, females (80%) were much more likely to feel legislation alone is not enough compared to males (67%). A similar perception was shared by those who identified as Hindu (83%). On the other hand, the level of agreement among non-heterosexual participants (52%) was lower than the average. In addition, those under 30 years old (14%) were more likely than average to disagree with the statement.

On the other hand, more than half of participants (55%) disagreed with the statement that people should be entitled to say whatever they want online. A quarter neither agreed nor disagreed. The level of disagreement was much higher among females (65%) compared to males (44%). A large majority of older adults and senior participants aged 50 and older (about 7 in 10) were against this idea as well. Meanwhile, the lowest level of disagreement was reported by young people, those aged 18-29 (33%) as well as Pacific participants (34%). On the other hand, among those who agreed with the argument that people should be entitled to say whatever they want online (16%), it was far more likely for males (25%) to agree with the argument compared to females (8%). Also 1 in 4 in the younger adult groups (those aged 18-29 and 30-39) indicated their agreement with the statement. Non-heterosexual respondents (36%)

were also more likely than average to agree with the statement.

Views were to some extent split when participants were asked whether people should just ignore any hateful online content targeted at them. While 39% disagreed with this statement, 34% agreed, with 21% indicating they neither agree nor disagree. Among those who disagreed, females (45%), those aged 70 and older (58%), Muslims (46%) and participants with disabilities (47%) reported a higher level of disagreement than average. On the other hand, regarding those who felt that people should just ignore any hateful online content targeted at them, this was higher among males (41%), Pacific (43%) and Asian respondents (42%) as well as those whose religious affiliation is Hinduism (58%).

A large majority of respondents (80%) agreed with the last statement in the question: "I believe that everyone has a role in addressing hateful online content". Participants aged 60-69 (90%) and 40-49 (88%) were more likely than average to agree. In the same way, higher levels of agreement with the statement were reported by those who identified as Christian (86%), and non-heterosexual participants (84%). In contrast, compared to the average, the lowest level of agreement with the statement was reported by Hindus (62%), people with disabilities (69%), and 30-39-year-olds (71%). Interestingly, Hindus (26%) and people with disabilities (8%) were more likely than average to disagree with the statement.

---

## PERCEPTIONS ABOUT ONLINE HATE SPEECH IN THE LAST YEAR

The last question of the survey asked all participants whether their feelings about online hate had changed in the last 12 months. This was a new question added to this year's survey. Data were collected through an open-ended question to allow answers to reflect participants' own knowledge, feelings, and understanding of online hate. Their answers were coded and grouped in themes. Some participants' comments reflecting the issues mentioned are included in this section for illustration.

## Participants' comments

*"Yes absolutely, people are just using these platforms to attack others in ways that is so damaging and leads people over the edge, some committing suicides, violence, shootings. Like the mosques in Christchurch, which puts NZ on the map for all the wrong reasons. It's devastating. Very sad to watch."*

**Male, 50-59 years, Samoan, identifies as gay/lesbian, targeted by online hate speech due to their sexual orientation.**

*"I'm less affected by it, by my own way of personally keeping myself safe, I make well thought-out judgement calls, I choose to be surrounded by positive environments, and make use of features on websites that block unwanted content ... we need to revolutionise the culture around the internet into a safe place for the generation of children growing up now."*

**Female, 18-29 years, NZ European, Māori, targeted by online hate speech due to their nationality, appearance, political views, ethnicity, gender, religion, and race.**

*"Yes. In light of Christchurch, but also of the view that the incident was not isolated in the sense that there is a current of racism and social grouping in NZ society that is felt by minority groups other than just Muslim – for example by Pacifica and in the reporting of 'South Auckland' as less than other areas of Auckland."*

**Male, 40-49 years, NZ European, Samoan, Dutch, targeted by online hate speech due to their appearance, political views, and ethnicity.**

*"The Christchurch shooting brought home the degree to which online content can be used and disseminated. I feel very concerned that there is little restriction on whom can actually access hate type material."*

**Male, 50-59 years, Māori, New Zealand European.**

*"i used to ignore them previously but recently it was all over the media so i had to listen to it and it was very disturbing and affected my emotional status."*

**Female 40-49 years, other ethnicity. Does not know if she was targeted.**

Most comments depict a change in participants' views of online hate speech in the last year. While several participants gave short "yes" answers, others expanded on their responses and explicitly indicated that the change in their feelings was due to the events around the Christchurch attack. Some commented on the need to "take action" which broadly included suggestions such as more education, sanctions, and/or law change. Similarly, others directed their call for action to stop hate speech towards social media platforms and other technology companies. Another theme that emerged from the qualitative data was the sense of being "more aware" about the pervasiveness of hateful speech in New Zealand, and the need to protect children and young people, in particular. Also, some participants commented that their views have changed although they have never been targeted or seen online hate, while others said they are "more cautious now" of what they do online in order to avoid this sort of content.

On the other hand, most participants who said their views about hate speech have not changed preferred not to expand further on their thoughts. However, a few indicated that the issues related to hate speech were "overstated" or in conflict with the notion of "free speech". A few also indicated that their views remain unchanged as they had already been opposed to any hateful content posted online.

## Concluding remarks

This report has presented trends in personal experiences of and exposure to online hate speech among adult New Zealanders based on nationally representative data. The findings from this study have also been compared with results from a similar research study conducted by Netsafe in 2018. In addition, this report explored people's perceptions about other issues related to hate speech. The goal of this study was to fill some knowledge gaps regarding the extent and nature of hate speech and provide reliable and robust evidence to inform public conversation on the topic.

## Participants' comments

*"Yes. They have changed dramatically with recent events in NZ and also certain discriminatory posts e.g. from sports players regarding sexual orientation."*

**Female, 30-39 years, Māori, New Zealand European, targeted by online hate speech due to their nationality and political views.**

*"I feel like people use the anonymity of the internet and the fact that it can be difficult to monitor to avoid being accountable for what they say or do."*

**Female 40-49 years, other ethnicity, targeted by online hate speech due to her political views.**

*"Yes. due to the events in Christchurch I feel that social media play a larger part in racial and religious (among others) hate that I do not know about due to the algorithms they run. This means that there are undercurrents in our society that can propagate where the mass of people are unaware."*

**Male, 30-39 years, Māori, not targeted.**

*"People say things online that they wouldn't say to someone face to face ... The problem is getting very bad in NZ especially people targeting other people race but then they say they are not racist. When confronted other people jump to their defence saying they are allowed to say what they want. This happens on many Facebook community groups everyday."*

**Male, 30-39 years, Asian, targeted by online hate speech due to his race and ethnicity.**

*"No. I think the media overstates the amount of "hate" there is in NZ and sensationalises stories to claim racism for comments that are not actually racist, I also think a lot of people are claiming to be targets of hate speech because it gets them attention rather than because they have actually suffered actual harm. I would be concerned about the definition of 'hate speech' and the conflict with freedom of speech."*

**Female, 30-39 years, NZ European, Pacific, does not know if she was targeted.**

One interesting finding involves the incidence of personal experiences of online hate speech. Compared to 2018, our data show a rise in the rate of people who reported being the target of hateful speech, from 11% to 15%. Statistically, this appears to be a slight increase; however, a closer look at the data in terms of key demographic variables such as age, gender, ethnicity and religion provides a better understanding of who exactly is more likely to be the target and how this seems to have shifted since 2018.

In this respect, for example, when looking at the results in terms of religious affiliation, the findings show that just over half of Muslims were targeted with hateful speech (52%) in the last 12 months, a rate far higher by 37 percentage points than the average population surveyed – followed by a third of Hindus. A similar pattern of higher prevalence of online hate speech is reported by members of minority ethnic groups (especially those who identified as an “other ethnicity”), people with disabilities, and non-heterosexual people (e.g. gay, lesbian, bisexual). Overall, these results are consistent with our 2018 study (Pacheco & Melhuish, 2018b) and support the argument that online hate speech is more likely to be directed towards minorities and/or vulnerable groups.

On the other hand, rather than being static, online hate speech seems to be a dynamic phenomenon. In this respect, when comparing our results with last year’s research, some interesting trends can be observed. For example, our study found that the incidence of online hate victimisation among older adults and seniors (aged 50 and over) significantly increased compared to the results from 2018. Similarly, there are variations in the reported experiences of online hate among minority ethnic groups. For those who identified themselves as an “other ethnicity” there was also an increase of personal experiences of hate speech compared to 2018, from 14% to 22%. In contrast, for Asian participants the prevalence in 2019 (11%) was lower than last year (16%). What is more, perceived reasons for hate speech victimisation similarly show fluctuations. For example, “appearance” was

among the top perceived reasons for receiving hateful content online in 2018; however, this year it was among the less common reasons. In addition, views that gender was the motivation behind online hate rose by 8 percentage points compared to last year. While our study did not investigate drivers explaining these variances in online hate speech victimisation in the last couple of years, it sheds new light on the changing nature of this phenomenon. It also highlights the need to keep measuring annual trends as a course of policy and research action for understanding and addressing it effectively in the aftermath of the Christchurch shooting.

Additionally, an important finding relates to the time the online hate incidents occurred in the previous year. As noted previously, fieldwork for this study was carried out during June 2019. Aggregated results from participants who were targeted with online hate shows that a third of the incidents (34%) happened in the last two months prior to completing the survey, just after the Christchurch attack. This reflects a substantially higher prevalence occurring within this short time period. The finding is also in line with our operational experience as Netsafe’s Call Centre registered a spike in the number of people reporting hateful speech incidents around the same time. While further research is required, the insight suggests that in the aftermath of a sensitive social or political event, such as the Christchurch shooting, the occurrence of online hate victimisation can, paradoxically, increase despite the unequivocal and widespread public condemnation of the attack.

In relation to exposure to online hate speech, our results show that about 3 in 10 have seen hateful content targeting someone else, a similar result to our study last year. As data gathering for this study was conducted just shortly after the Christchurch attacks, we initially assumed that exposure was likely to increase – based on the incidents reported to Netsafe’s contact centre team soon after the shooting, as initially mentioned in this report. Also, previous research on terrorist attacks in London and Paris suggest that hate speech exposure increases around certain social/political events

and under conditions of fear, uncertainty, and polarisation (Kaakinen, Oksanen, & Räsänen, 2018; Williams & Burnap, 2016). However, this does not appear to be the case in our study. A possible explanation for this might be in the methodological approaches underpinning overseas studies. Rather than gathering nationally representative data, these two studies analysed posts from one specific social media platform or sampled only teens and young adults. Another possible explanation for this is that making it illegal to have or share any content related to the Christchurch shooting in New Zealand might have helped to prevent the potential increase of exposure to online hate in the country in the aftermath of the attack.

Another key contribution of our study is the exploration of adult New Zealanders' perceptions of online hate-related issues. In this respect, our findings reveal that a significant proportion of New Zealand adults think that online hate is spreading. This finding is supported by the qualitative data collected for this study which also depict people's concerns about the growth of this phenomenon. What is equally interesting is that while most New Zealanders support new legislation to counteract hateful speech, they also believe that legal measures are not enough. This point has implications for policy and practice as it suggests that the introduction of any legal mechanism might need to be accompanied by multi-faceted policy interventions such as rights awareness campaigns or educational activities focusing on children, young people and other vulnerable groups, for instance. As our results also show that most New Zealanders believe that everyone plays a role in addressing hateful content, achieving public support and engagement with more comprehensive online hate responses and strategies seems plausible and feasible.

In addition, public perceptions regarding the role of social media platforms are noteworthy. As previously described, a large majority of New Zealand adults, over 8 in 10, think that social media companies should do more to halt the

spread of online hate. This reflects a stronger position compared to public views regarding specific actions such as platforms stopping livestreaming<sup>9</sup>. This finding has implications for social media companies as it presents an opportunity to work closely and openly with governments, their users, and other interested parties in tackling the spread of abusive content. This includes exploring varied approaches to complement the use of machine learning and computational methods to detect and remove hateful content from their services.

Finally, despite the exploratory nature of this study, we are confident in the relevance of its findings to inform public conversation, contribute to the growing body of international research, and help us all reflect critically on the nature, extent, and impact of online hate speech on the basis of robust research evidence.

## What's next?

Netsafe along with Australia's eSafety Commissioner and the UK's Safer Internet Centre are working on a research report which will include comparative data on online hate speech. The report will be released at the start of 2020.

To contact Netsafe for more information about its research programme email [research@netsafe.org.nz](mailto:research@netsafe.org.nz)

For more information about New Zealand's Harmful Digital Communications Act 2015 and Netsafe's Approved Agency role visit: <https://www.netsafe.org.nz/hdc-act/>

If you are experiencing online abuse or harassment or another online issue, Netsafe has a free helpline for people in New Zealand. The helpline is open from 8am-8pm Monday to Friday, and 9am-5pm on weekends. Contact Netsafe toll free on 0508 NETSAFE, by emailing [help@netsafe.org.nz](mailto:help@netsafe.org.nz) or visiting <https://netsafe.org.nz/report>

---

<sup>9</sup> See <https://www.stuff.co.nz/national/christchurch-shooting/112695810/most-kiwis-want-livestreaming-halted-until-violent-content-can-be-curbed-survey>

## References

- ActionStation. (2019). The people's report on online hate, harassment and abuse. Retrieved 11 December 2019, from <https://peoplesharassmentreport.com/>
- Greenfield, C., & Menon, P. (2019). Before mosque attacks, New Zealand failed to record hate crimes for years. Retrieved 11 December 2019, from <https://www.reuters.com/article/us-newzealand-shooting-intelligence-anal/before-mosque-attacks-new-zealand-failed-to-record-hate-crimes-for-years-idUSKCN1RB0PW>
- Kaakinen, M., Oksanen, A., & Räsänen, P. (2018). Did the risk of exposure to online hate increase after the November 2015 Paris attacks? A group relations approach. *Computers in Human Behavior*, *78*, 90–97. <https://doi.org/10.1016/J.CHB.2017.09.022>
- Pacheco, E., & Melhuish, N. (2018a). Harmful digital communications in New Zealand: Annual Population Survey 2017. <https://doi.org/10.2139/ssrn.3128121>
- Pacheco, E., & Melhuish, N. (2018b). Online hate speech: A survey on personal experiences and exposure among adult New Zealanders. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3272148>
- Pacheco, E., & Melhuish, N. (2019). *New Zealand children's experiences of online risks and perceptions of harm. Evidence from Ngā taiohi matihiko o Aotearoa - New Zealand Kids Online*. Wellington, New Zealand. Manuscript in preparation.
- Spoonley, P. (2018). *Te Whakamauāhara ki te Ipurangi: Hate speech in the age of the internet*. Palmerston North: Massey University.
- Williams, M. L., & Burnap, P. (2016). Cyberhate on social media in the aftermath of Woolwich: A case study in computational criminology and big data. *British Journal of Criminology*, *56*(2), 211–238. <https://doi.org/10.1093/bjc/azv059>

Measuring trends in online hate speech victimisation and exposure, and attitudes in New Zealand

Wellington, New Zealand, December 2019

[www.netsafe.org.nz](http://www.netsafe.org.nz)

[research@netsafe.org.nz](mailto:research@netsafe.org.nz)

Retrieved from: <https://www.netsafe.org.nz/online-hate-speech-report-2019>

ISBN: 978-0-473-50802-9



**ATTRIBUTION-NONCOMMERCIAL-SHAREALIKE**

<https://creativecommons.org/licenses/by-nc-sa/4.0/> [English]

<https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode.mi> [Te Reo Māori]