netsafe

Online Hate Speech

A survey on personal experiences and exposure among adult New Zealanders

Prepared by Dr. Edgar Pacheco and Neil Melhuish

Foreword

In 2015 the New Zealand government passed the Harmful Digital Communications Act[1] to deter, prevent and mitigate harm caused by digital communications and to provide victims of harmful digital communications with a quick and efficient means of redress. At the heart of the Act are ten communications principles that together describe a broad range of challenging online communications that people can send and receive. We have drawn on principle 10 as the focus for this study. It says that a digital communication should not:-

> denigrate a person's colour, race, ethnic or national origins, religion, gender, sexual orientation or disability

This principle extends to a limited set of communications which could be considered hateful in nature. We wanted to better understand the experiences of New Zealanders receiving, or being exposed to, these kinds of communications. This is important for us - in our Approved Agency role under the Act – so we can continue to provide the right kind of advice and support for those that need it. We are also motivated by the idea that harmful digital communications are a shared problem, requiring a shared solution involving government, industry, academia, community groups, schools, families and whānau. Simply, at Netsafe we believe that reliable evidence based on New Zealanders' experience is vital if our individual or collective efforts are to be effective.

Online hate speech is not a new topic. So, it was a surprise to find that, to the best of our knowledge, this is the first study to look at the topic of online hate speech in New Zealand. It is also timely, as the topic of online hate seems to have emerged into in our collective consciousness over the last couple of years, delivered by a series of international and New Zealand media stories. This study reaches beyond the headlines to ask New Zealanders' about their experiences of online hate speech, including why they think they received it, and its impact on them.

What we have found raises important questions about online hate speech in New Zealand.

The findings in this report reveal how some groups of New Zealander's are more likely to be targeted by online hate speech than others, depending on factors such as ethnicity, religion, disabilities, sexuality and gender. Worryingly, there are similarities with findings from other studies we have carried out into harmful digital communications in NZ.

We also asked participants in the study to tell us in their own words about the impact that receiving such digital communications had on them. Many people shared online experiences they found difficult at the time and that had a lasting impact on them. Their comments provide a brief insight into the myriad of personal stories that sit underneath the statistics. They also highlight the many challenges that remain in tackling issues such as online hate.

We believe that the insights in this report should represent an important step forward in our understanding of the issues surrounding this topic.

---

[1] For further information see http://www.legislation.govt.nz/act/public/2015/0063/latest/whole.html

Executive Summary

The purpose of this report is to present findings from an exploratory study regarding adult New Zealanders' personal experiences of and exposure to online hate speech. The study, which is part of a larger project regarding online risks and harm, is the first of its kind in the country. Netsafe, the Approved Agency under the Harmful Digital Communications Act 2015, carried out this study to contribute research evidence and insights to the understanding of the extent and impact of online hate speech in New Zealand.

Online hate speech[2] has been a topic of public concern and research interest for some time. Initially the focus of this centred on the proliferation of online groups and websites promoting and distributing discriminatory content. Since the introduction of more interactive tools and platforms in the mid-2000s that enabled new and faster ways of disseminating content in a relatively anonymous fashion, concerns about online hate speech becoming a pervasive behaviour have increased.

Current research and analysis acknowledge the complex nature of online hate, the mediating role of technology and the influence of other contextual factors (e.g. social or political events). However, despite the growing attention on the topic, New Zealand-based research looking at personal experiences and/or exposure to online hate is surprisingly absent. This study seeks to address this gap. It builds on existing international research on young people's experiences to explore those of the adult New Zealand population based on a nationally representative sample.

The research instrument used for this study was an online survey. The survey was administered by Colmar Brunton in June 2018. The maximum margin of error for the whole population is ±3.1% at the 95% confidence level. The sample is representative of the wider population on key demographics: age, gender, ethnicity, and location.

In looking at the extent and frequency of personal experiences, this study found that online hate speech is more prevalent among minority groups. It was equally interesting to find that male and younger adults were more commonly targeted than females and older adults. The report also provides insights into the most common reasons participants perceived for being targeted with hateful content online. Furthermore, it reports the negative impact that online hate had on people who personally experienced being targeted by this content. The insights about exposure to online hate reflect results from similar international research.

This report provides the starting point to better understand the extent and impact of online hate speech in New Zealand. We believe this study is also a useful source that adds to the growing international body of research on this topic. Finally, the delivery of this report reflects Netsafe's commitment to providing the online safety community with research-based insights about online behaviours and the impact of digital communications in New Zealand.

---

[2] A working definition of online hate speech is provided in the Methodology section.

netsafe

Key findings

PERSONAL EXPERIENCES OF ONLINE HATE SPEECH

- Overall, 11% of New Zealand adults reported to have been personally targeted with online hate speech in the prior year.

- Online hate[3] was more prevalent among:
  - Minority ethnic groups, particularly Asians, followed by those who identified themselves within the 'other' ethnicity category, then Māori, and Pacific participants.
  - Males (13%) compared to females (8%).
  - Younger adults, especially those between 18 and 39 years old.
  - People with disabilities (15%) compared to those without impairments (10%).
  - Non-heterosexual respondents (e.g. gay, lesbian).

- Of those targeted, about 6 in 10 reported a negative impact from the experience. Most reported being affected emotionally but also exhibiting changes in their behaviour. A third indicated not being affected.

- 'Religion' was the most frequent perceived reason for being personally targeted with online hate speech. This was followed by political views, appearance, race, and ethnicity.

- In relation to gender:
  - Males believed they were targeted mainly because of their religion, race, political views, and ethnicity.
  - Females, meanwhile, concurred on religion but rated gender and age significantly higher than males.

EXPOSURE TO ONLINE HATE SPEECH

- In the prior year, about 3 in 10 of all participants had seen or encountered online hate speech that targeted someone else.

- Exposure to online hate was more common among participants within 'other' ethnic groups.

- Those who were exposed to online hate believed that people were more commonly targeted because of their religion, race and/or ethnicity.

- In the prior year, 1 in 20 had intentionally visited a website, online forum and/or social media site that promotes online hate speech.

- Intentional engagement with these online sites was more common among males, young people under 30 years old, Asians, and those who identified themselves as non-heterosexual.

---

[3] Note we use the terms 'online hate speech' and 'online hate' interchangeably.

netsafe

Contents

List of Figures and Tables

netsafe

Introduction

Last year Netsafe conducted the first Annual Population Survey in the context of the Harmful Digital Communications Act 2015. The purpose was to gauge adult New Zealanders' experiences with unwanted digital communications[4], and to develop a better understanding of digital risks and their potential harm. At that time, around a third of participants reported having experienced one or more than one type of unwanted content online in the prior year. Interestingly, one of the most common experiences among participants (9%) was to receive a digital communication(s) that said offensive things about their lifestyle or their religious or political beliefs. The report for that study was published early this year (see Netsafe, 2018a). With this valuable but limited insight, Netsafe decided to explore in more detail the extent and impact of online hate speech. Our interest was also triggered by the lack of evidence on the topic as agencies in New Zealand are not required to systematically collect data on hate speech (Spoonley, 2018).

This section of the report presents a summary of relevant literature focusing on available statistical research on personal experiences and exposure to online hate speech. The purpose is to provide a context for the study and support interpretation and comparison of our results whenever possible.

Background

Online hate speech is a complex phenomenon

In addition to the multiple opportunities afforded by digital technologies they also provide a means for spreading hateful content. Accounts of online hate activities date from the mid-1980s when a Commodore 64 desktop computer with a telephone modem connection was used to allow skinheads, Klansmen, and Neo-Nazis to communicate and download electronic bulletin boards (Duffy, 2003). Since then, initial research interest on online hate centred on the growth of hateful websites and the characteristics and dynamics of online hate groups, from white supremacists to terrorist groups, but also on the type and the persuasiveness of the messages disseminated (see Burris, Smith, & Strahm, 2000; Duffy, 2003; Gerstenfeld, Grant, & Chiang, 2003; Lee & Leets, 2002).

The inception of more innovative tools, such as social media platforms, has given rise to new challenges and concerns in regard to online hate (Gagliardone, Gal, Alves, & Martinez, 2015; Keipi, Näsi, Oksanen, & Räsänen, 2016; Silva, Mondal, Correa, Benevenuto, & Weber, 2016). In this context, as commentators highlight (see Brown, 2017; Gagliardone et al., 2015), online hate speech is low cost, can be facilitated anonymously and pseudonymously, is easy to access, is instantaneous, can reach a larger audience, and can be spread via different formats across multiple platforms. It also raises cross-jurisdictional issues in regard to legal mechanisms for combatting it.

However, one of the long-term challenges has been the lack of an agreed understanding and definition of online hate speech (Alba, 2017; Spoonley, 2018). This has methodological implications for researchers, affecting the way data are collected, interpreted, and reported.

---

[4] For a definition of unwanted digital communications see the Glossary section in this report.

Similarly, as Alba (2017) describes, it creates practical difficulties in terms of preventing or removing hateful content online.

Equally important are the evolving perceptions of what constitutes online hate speech. For example, legal perspectives initially focused on expressions of racism and xenophobia online but research has also moved towards instances of online hate in relation to gender, disability, and sexual orientation (Gagliardone et al., 2015; Silva et al., 2016). In regard to the latter point, HateBase, a crowdsourced database that relies on Artificial Intelligence (AI) to track hate speech in social media, has created a list of the most frequent types of hate speech based on the collection of the range and number of words used in abusive content (HateBase, 2018). According to this list, at the time of writing, online hate speech mainly targets ethnicity, then nationality, and class. These are followed by religion, sexual orientation as well as gender and disability. These different forms of online hate are also reflected in the policies and community standards of major online platforms such as YouTube and Facebook (see Facebook, n.d.; YouTube, n.d.). For this study, we provide a working definition of online hate speech which is included in the Methodology section of the report.

Evidence shows that more instances of online hate speech occur on social media (Hawdon, Oksanen, & Räsänen, 2014; Keipi et al., 2016; Mondal, Silva, & Benevenuto, 2017). This has attracted attention in exploring online hate on specific social media platforms (see Miškolci, Kováčová, & Rigová, 2018; Williams & Burnap, 2016) and/or using computational methodologies to detect, remove, and understand the dynamics of hateful content distributed through these tools (Mondal et al., 2017).

Major social media platforms are also using AI and computational methods to detect and remove online hate speech from their services. However, there are concerns about the bias of computer-based assessments of what online content is acceptable (Alba, 2017) or how the learning of an AI system can be used in unintended and unanticipated ways (Buranyi, 2017). Another response to mitigating online hate is the introduction of regulatory frameworks, such as Germany's recent law that demands social media platforms remove hateful content from their services within 24 hours (Martin & Rolph, 2016). In addition to computational and legal responses, online hate is argued to be effectively addressed through education that helps people to be ethically-reflective users, develop their digital literacy skills (Gagliardone et al., 2015) and their ability to respond through counter-speech (Martin & Rolph, 2016). However, despite the good intentions, rigorous assessment of the implementation and effectiveness of these different approaches to prevent and mitigate the impact of online hate is still lacking (Blaya, 2018).

What the evidence on personal experiences and exposure tells us

Available international research on personal experiences of and exposure to online hate speech has primarily looked at children and young people as well as the content, dissemination and legal consequences of hateful material (Oksanen, Hawdon, Holkeri, Näsi, & Räsänen, 2014). Additionally, in most cases, measures vary not only across countries but also within a single country due to research design and other factors.

In the context of children's experiences, a representative study in the United States (Ybarra, Mitchell, & Korchmaros, 2011) among participants aged 10-15 years old found that only a small percentage of children visited hate sites in 2006 (2.6%), 2007 (2.3%), and 2008 (3.5%). These findings contrast with longitudinal research that suggests other forms of online risks and

netsafe

challenges such as online harassment have steadily increased in the United States (see Jones, Mitchell, & Finkelhor, 2013). In Europe, another study in 25 countries (Livingstone, Haddon, Görzig, & Ólafsson, 2011) revealed that 12% of European children (11-16 years old) encountered hate messages online in the prior year, adding that exposure to this sort of material increases with the age of the child. In Australia, a recent study commissioned by the eSafety Commissioner (2017) found that 56% of children and teens aged 12-17 years old have seen racist comments online and that 53% have seen or heard hateful comments about cultural or religious groups. In New Zealand evidence about the prevalence of experiences and exposure to online hate speech among children is yet to be gathered.

Some attention has also been paid to youths and young adults' personal experiences of and exposure to online hate speech. Among the few available studies, Hawdon, Oksanen and Räsänen's (2014) online survey with young Americans aged 15 to 30 years old found that some 53.4% of respondents had been exposed to online hate in the prior three months. In addition, 15.8% said they had been personally targeted. The study also found that young people reported that the hateful content they were exposed to was most commonly focused on people's sexual orientation and ethnicity/nationality followed by their political views, religious convictions, gender and physical appearance. Similarly, Costello, Hawdon, Ratliff and Grantham (2016; 2017) looked at young Americans but broadened the age group (15 to 36 year olds). In one of the studies, when asked whether they have been personally targeted with online hate at any time of their life, 23% of respondents said "yes". These participants also indicated that the most common reason for receiving this sort of content was ethnicity or race and then appearance, religious beliefs, sexual orientation, political views, nationality, and sex/gender (Costello et al., 2017). Costello and his colleagues also asked young Americans about online hate exposure. In this respect, a significant 65.4% indicated having come across hateful content during the three months prior to completing the survey (Costello et al., 2016).

There have also been other attempts to measure exposure to online hate across countries. A study on young people found that exposure was a common experience but varied between the countries surveyed: Finland, Germany, the United Kingdom, and the United States (Hawdon, Oksanen, & Räsänen, 2017). Conducted between 2013 and 2014, the study asked young participants whether they have seen hateful content online in the last three months. The findings revealed that exposure was higher among American respondents (53%) followed by Finns (48%), Britons (39%) and Germans (31%). Regarding the estimated likelihood of being exposed, there were some differences by country with Americans and Finns having a higher chance of being exposed to online hate. The study also found that in the United States and the United Kingdom online hate exposure increased with young people's age. Similarly, while there were significant differences across countries, young male Americans were more likely to be exposed to online hate compared to female Americans.

On the other hand, evidence also suggests that the increase in exposure to online hate speech can be associated with the nature and impact of certain social events and conditions. For instance, Williams and Burnap's (2016) study explored the impact that a 2013 terrorist attack perpetrated by Islamist extremists in London had among social media users in the United Kingdom. The authors collected data from Twitter for a two-week period after the attack. They found that the attack triggered the production and propagation of online hate on Twitter targeting religion and race. Recently, Kaakinen, Oksanen and Räsänen (2018) looked at the impact of terrorist attacks in Paris in November 2015. For this study they compared two datasets from Finnish young people aged 15 to 30: one collected in 2013 and the other in 2015, one month after the attacks. The findings supported their hypothesis that online hate

exposure increases in a "social climate of fear, uncertainty, and polarization" (p. 90) as around 47% of respondents had encountered online hate during 2013, while 74% did so in 2015.

As described above, over the years research interest in online hate speech has moved from the proliferation of hateful websites and online groups to the prevalence of exposure, particularly among children and young adults, and the implementation of approaches to mitigate online hate. However, in reviewing the literature, we have also found that New Zealand lacks statistical insights about this issue. This study tries to address the knowledge gap by looking at one specific aspect: the self-reported experiences of online hate among adult New Zealanders in general. The next section of this report describes the methodological approach that guided this study.

Methodology

Currently, there is a lack of statistical insights about online hate speech in New Zealand. For this reason, we adopted an approach that mainly relied on numerical data. Our approach was pertinent and timely to support current national discussion on the topic. The objectives of the study were to explore the self-reported experiences of New Zealand adults regarding online hate speech and their exposure to this type of content. To the best of our knowledge, this study is the first of its kind in the country.

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?
- What is the extent of exposure to online hate speech among adult New Zealanders?

As described in the Background section, there is no consensus on the definition of online hate speech. However, in order to answer the research questions, we used a working definition based on previous social and behavioural research on the topic, which has also applied quantitative methodology, particularly surveys, as the technique for data collection. In doing so, we sought to obtain results that, to some extent, can be compared with those informing our study. In this sense, we define online hate speech 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 (Kaakinen et al., 2018; Schmidt & Wiegand, 2017).*

In the following sub-sections, we describe the data collection technique used for the study, the characteristics of the research sample, aspects related to research ethics, and the limitations of the study.

Survey tool

The instrument for data collection was an online survey. The use of online surveys for government, academic, and commercial research is widespread. The advantage of using this type of research instrument, compared to traditional paper-based surveys, is that it allows access to a larger sample, incurs lower administrative costs, and provides higher flexibility as well as greater efficiency (Evans & Mathur, 2005). Despite these benefits, a downside is that it is restricted to the online population (Bryman, 2008). However, this last point was not an issue for this study as we were interested in gathering data from adults that typically navigate online. Furthermore, recent research on online hate speech has successfully used online surveys as the means for data collection (see Costello et al., 2016, 2017, Hawdon et al., 2014, 2017).

The development of the research instrument for this study was informed by available international research on the topic (Costello et al., 2016; Hawdon et al., 2017; Kaakinen et al., 2018; Keipi et al., 2016). While most of these studies looked at online hate exposure, our survey also explored personal experiences of online hate speech. We also relied on Netsafe's operational experience with reported cases of online hate to refine survey questions and/or scaling responses. Similarly, we received feedback from experts on the field and key stakeholders regarding our research design. Data collection was administered by Colmar Brunton and took place during June 2018. The online survey was piloted before going live.

The survey questionnaire included six questions[5]. The first question measured participants' personal experiences of online hate speech in the last 12 months. We focused on the prevalence and frequency of these experiences. The scale of responses included: never, once, a few times (2-4), many times (5 or more), and I don't know. Then, those participants who said they had experienced online hate were asked a follow-up question regarding perceived reasons for being the target of hateful content. For this question the list of options included race, ethnicity, gender, nationality, and sexual orientation among others.

While we primarily sought to gather quantitative data, our online survey included an open-ended question that provided us with some qualitative data about the impact of online hate from the perspective of those who were targeted with it. We collected useful insights based on participants' own words and views, but, as we only included one open-ended question, the value of and need for further qualitative evidence in future research became apparent to us.

Subsequent survey questions measured exposure to online hate. Note that by exposure we meant seeing or encountering hateful content targeting someone else. In this respect, we asked all participants whether they have been exposed to hateful content in the last 12 months. We consistently applied the scale used to measure personal experiences, which, as previously described, included the following options: never, once, a few times (2-4), many times (5 or more), and I don't know. Then, participants who were exposed to online hate answered a question about the reasons they believed others were targeted with hateful content.

The last survey question sought to measure participants' engagement with online sites that promote or distribute hateful content. Specifically, we asked them to indicate whether they have intentionally visited such a website, online forum and/or social media group.

---

[5] See Appendix for six-question research instrument.

netsafe

Sample

The study collected data from New Zealanders aged 18+. A total of 1,001 participants completed the online survey. Colmar Brunton sampled participants from its online research panel, offering them incentives for participation. The use of survey panels is argued to increase the validity of responses as participants recruited from these panels tend to be more serious about answering survey questions (Wansink, 2001). The demographic details provided by the participants included information such as age, gender, ethnicity, and location.

Regarding gender, females represented 52% of the total sample while males made up the remaining 47.7%. Only 0.03% of participants identified as gender diverse. In terms of ethnicity our sample was distributed as follows: NZ European/Pākehā (71%), Māori (11%), Pacific (5%), Asian (11%), and other ethnicity (9%). Māori and Pacific sub-populations were deliberately 'boosted' beyond what would be achieved in a random sample to ensure representation.

Data about sexual orientation[6] was also collected. As there were relatively few respondents who identified as gay, lesbian, bisexual, or other, these respondents were grouped as non-heterosexual (n=62) to make statistically robust comparisons with heterosexual respondents. This approach is common in survey-based research.

Data collected from this representative sample allowed the analysis of sub-groups. The margin of error was +/- 3.1% at 95% confidence on total results.

Research ethics

As previously indicated, this research is part of a larger project that looks at New Zealanders' experiences of digital risks and harm. Because of the sensitive nature of the topics involved, we provided participants with pertinent information about the purposes of the study, asked them for their consent to participate, and guaranteed that their responses were confidential and that their data will be kept protected. Colmar Brunton, in coordination with Netsafe, also followed industry standards included in the Research Association's Code of Practice. The research instrument was piloted to identify whether there was any risk of causing distress to the respondents. 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 us.

---

[6] For a definition of sexual orientation and gender see the Glossary section in this report.

Limitations

While we have tried to follow the tenets that underpin rigour in social science research to explore the nature and impact of online hate speech in New Zealand, we are also aware of the limitations of our approach.

First, while our findings rely on representative data, our questionnaire was short, comprising only six questions due to practical issues. Therefore, the findings provide a broad picture of online hate speech in New Zealand. We have tried to manage this challenge by including an open-ended question, so participants could freely share their own views and expand further on the topic.

Second, we acknowledge that self-reported data are subject to respondents' personal perspectives, and that they might have different definitions of hate speech. Another limitation is that the study collected data from a specific point of time. As previously described, evidence suggests that the dissemination and, thus, exposure to hateful content online might be influenced by context (e.g. a terrorist attack, intense media coverage on a sensitive topic, a policy decision etc.). At the time of data collection, there were no specific controversial political or social events that generated public attention in New Zealand. This point needs to be considered in the interpretation and comparison of results.

netsafe

Findings

This section presents findings about adult New Zealanders' personal experiences of and exposure to online hate speech in the last 12 months. It also reports on perceived reasons for receiving hateful content, the impact of online hate, and participants' intentions to visit online environments (e.g. websites, online forums, social media) that target people because of their personal identity and/or beliefs. Differences in terms of key demographics such as age, gender, and ethnicity, are also described in this section.

Personal experiences and frequency of online hate speech

Highlights
- 1 in 10 adults reported to have personally experienced online hate speech in the prior year.
- Online hate mainly targets male, minority ethnic as well as younger age groups.
- There is indication of a higher rate among those identified as non-heterosexual.

Survey participants were asked 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. While a large majority (83%) answered negatively, 11% said they have experienced online hate speech in person one or more times in the last 12 months (Figure 1). Of those participants who answered positively, half experienced online hate speech 2-4 times.

YES
4% Once
5% A few times (2-4)
1% Many times (5+)
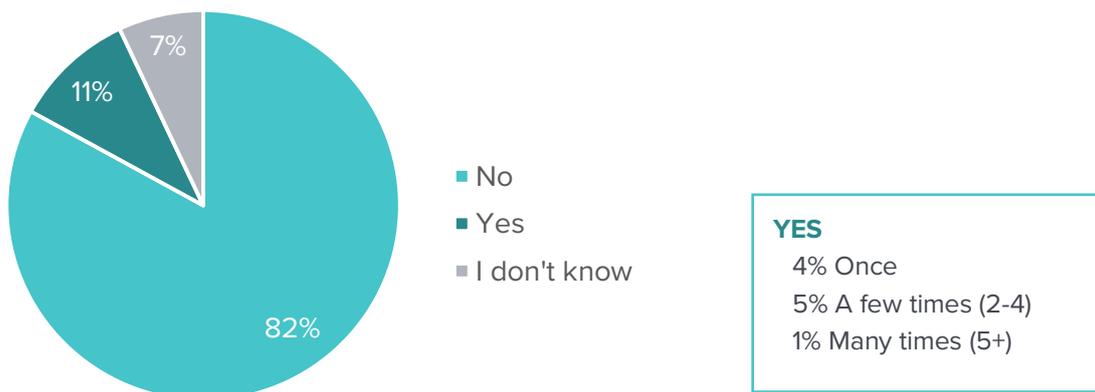
Figure 1. Overall personal experiences of online hate speech

Base: All respondents (1001).

Some relevant differences were identified when the data was disaggregated in terms of gender. In this respect, for male participants (13%) it was more common to be the target of online hate speech compared to females (8%). See Figure 2.

netsafe
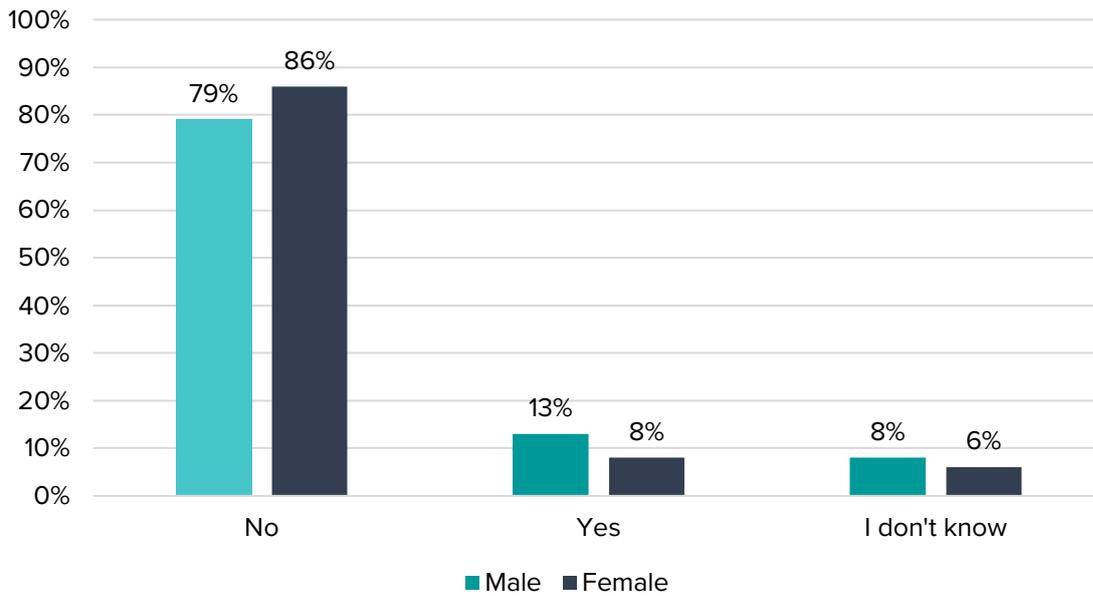
Figure 2. Personal experiences of online hate speech by gender

Base: All respondents (1001).

The frequency of online hate speech in the context of gender is also described below (see Figure 3). Experiencing online hate speech 'a few times' was higher among males than females.
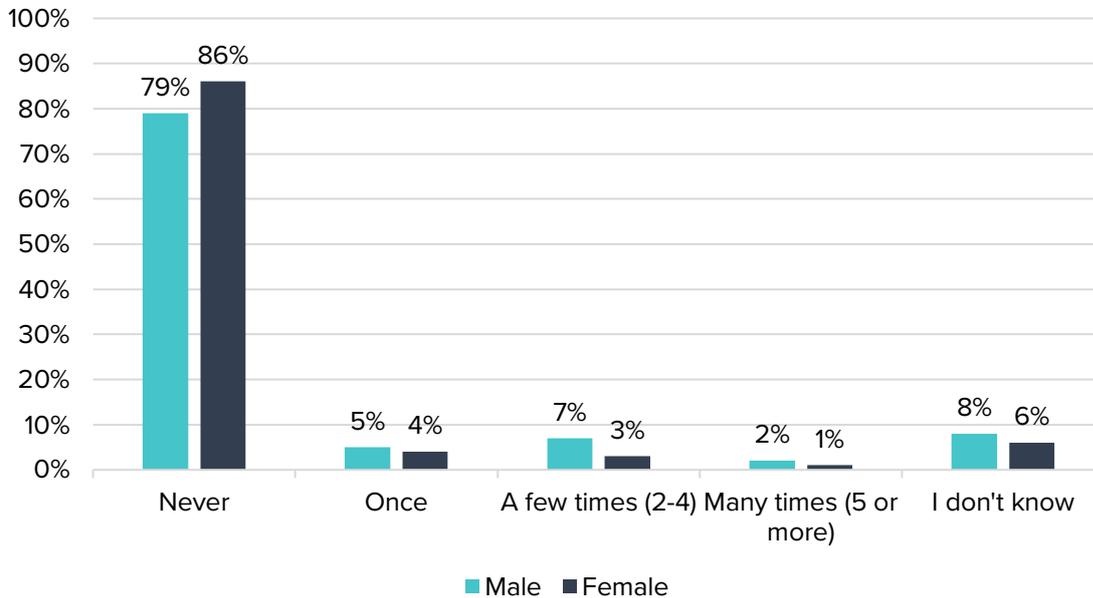
Figure 3. Frequency of personal experiences of online hate speech by gender

Base: All respondents (1001).

The data also shows that personal experiences of online hate speech were more common among minority ethnic groups. Aggregated data shows that 16% of Asian participants experienced online hate speech one or more times in the prior year. They were followed by those who identified themselves as an 'other' ethnicity (14%) and then Māori and Pacific

netsafe

participants, both at 13%. For NZ European/Pākehā (9%), experiences of online hate speech were less common. See Table 1.

Table 1. Personal experiences of online hate speech by ethnicity

Frequency	NZ European/ Pākehā	Māori	Pacific	Asian	Other
Never	86%	75%	77%	71%	76%
Once	4%	4%	7%	8%	2%
A few times (2-4)	4%	7%	5%	6%	10%
Many times (5 or more)	1%	2%	1%	2%	2%
I don't know	5%	12%	10%	13%	10%

Base: All respondents (1001).

Interestingly, as also evident in Table 1, the percentage of those who experienced online hate speech a few times was higher among participants with 'other' ethnic backgrounds (10%). Also, compared to other ethnic groups, for Asian and Pacific respondents it was more common to be the target of online hate only once: 8% and 7%, respectively.

In regard to age, experiences of online hate speech varied with participants in the younger age groups more commonly receiving this type of content (see Table 2). This was the case for those aged 18-29 and 30-39 years old: 16% and 18%, respectively.

Table 2. Personal experiences of online hate speech by age

Frequency	18 – 29 years old	30 – 39 years old	40 – 49 years old	50 – 59 years old	60 – 69 years old	70 years or older
Never	69%	75%	82%	88%	93%	97%
Once	7%	10%	3%	3%	1%	2%
A few times (2-4)	8%	5%	8%	4%	-	1%
Many times (5 or more)	1%	3%	1%	2%	-	-
I don't know	14%	8%	6%	3%	5%	-

Base: All respondents (1001).

Respondents with a disability who answered the online survey also indicated that they experienced online hate at a higher rate compared to non-disabled participants. As Figure 4 shows, 15% of these respondents were targeted with online hate compared to 10% of participants without impairments.

netsafe
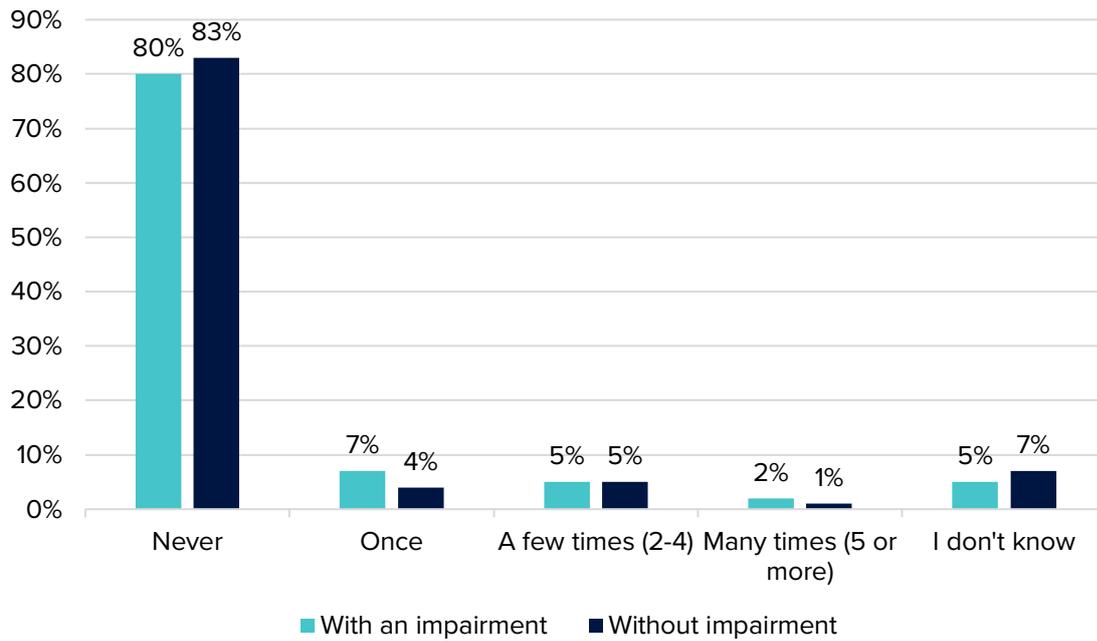
Figure 4. Personal experiences of online hate speech by disability

Base: All respondents (1001).

The online survey also gathered data about online hate in the context of participants' sexual orientation (see Table 3). About 9% of those who identified themselves as heterosexual indicated to have received hateful digital communications one or more times in the prior year. However, for non-heterosexual participants (e.g. gay, lesbian) the rate was higher, with 26% of them receiving such communications.

Table 3. Personal experiences of online hate speech by sexual orientation

Frequency	Heterosexual	Non-heterosexual
Never	86%	66%
Once	4%	8%
A few times (2-4)	4%	12%
Many times (5 or more)	1%	6%
I don't know	5%	7%

Base: All respondents (1001).

Note that these percentages are indicative only.

netsafe

Perceived reasons for receiving online hate speech

Highlights
- Religion followed by political views, appearance, race, and ethnicity were the most frequent reasons behind online hate speech.
- Males believed they were targeted mainly because of their religion, race, political views, and ethnicity.
- Females, meanwhile, concurred on religion but considered gender and age significantly higher than males.

Survey participants who reported receiving online hate in the prior year (n=112) were asked to indicate the reasons for being targeted by this kind of digital communication. Overall, religion was mentioned as the most frequent reason followed by appearance, political views and race. Ethnicity and nationality were also mentioned. For details see Figure 5.

| Reason	Percentage
Religion	24%
Appearance	20%
Political views	20%
Race	20%
Ethnicity	18%
Nationality	14%
Sexual orientation	10%
Gender	8%
Age	6%
Disability	5%
Don't know	8%
Other	13%
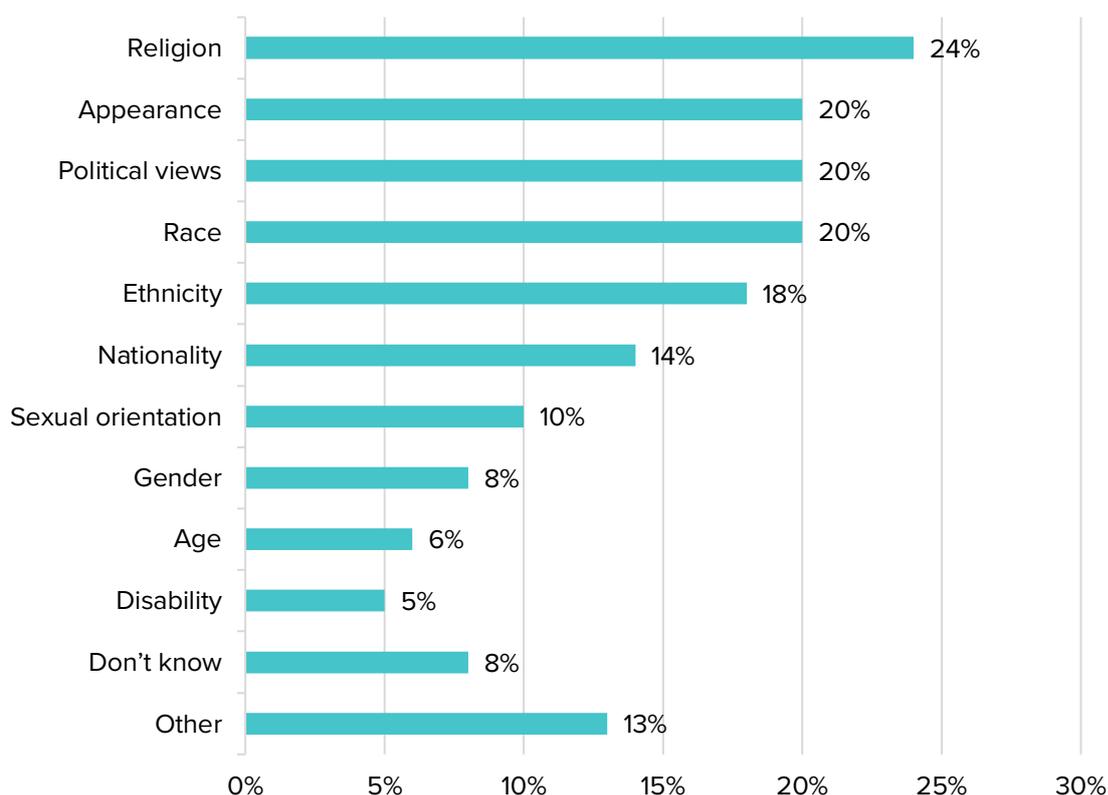
Figure 5. Perceived reasons for receiving online hate speech

Base: Respondents who had experienced online hate speech in the last 12 months (112). Note multiple responses were allowed.

netsafe

In terms of gender, male and female participants considered 'religion' to be the most frequent reason for receiving online hate speech. However, these two groups weighted other motives differently. As Figure 6 describes, for male participants ethnicity, political views and race were, for example, some of the most frequent reasons behind their personal experiences of online hate speech. On the other hand, responses regarding gender and age as motivations for receiving hateful digital communications were rated significantly higher by females compared to males.
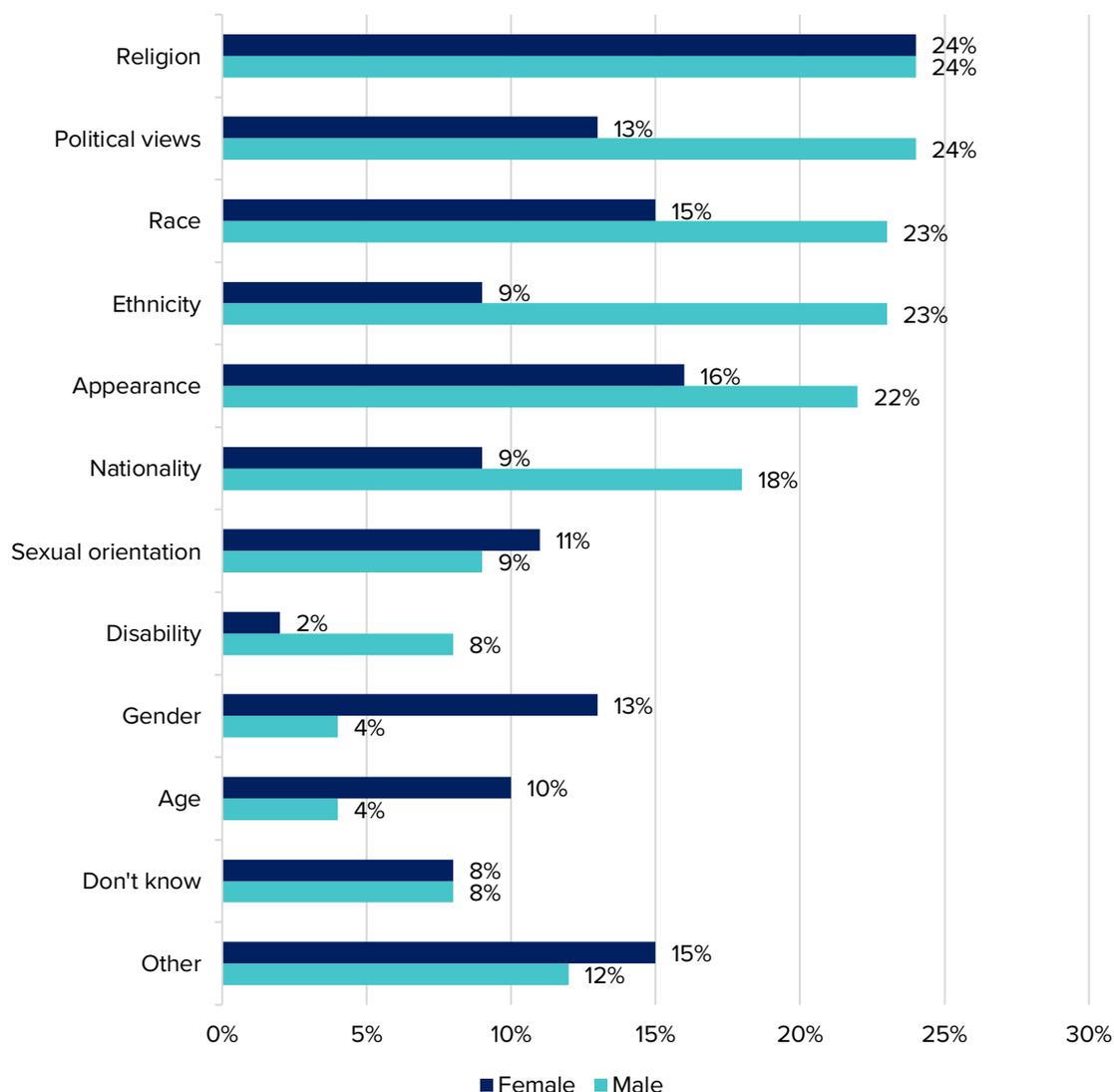
Figure 6. Perceived reasons for receiving online hate speech by gender

Base: Respondents who had experienced online hate speech in the last 12 months (112). Note multiple responses were allowed.

In terms of ethnicity, most participants from minority ethnic groups concurred that their ethnicity was the reason for being targeted with hateful content online. Participants of Asian backgrounds mentioned that their ethnicity (42%) and religion (28%) were the most frequent reasons for receiving online hate. Meanwhile, those who identified themselves as part of 'other' ethnic groups said that motivations based on their race (44%) and ethnicity (42%) were used to target them. For their part, Māori also highlighted ethnicity (36%) and race (33%) as motivations. In contrast, Pacific respondents pointed out race (31%) and religion (26%), while

netsafe

NZ European/ Pākehā participants indicated religion (26%) and political views (22%). See more details in Table 4.

Table 4. Perceived reasons for receiving online hate speech by ethnicity

Perceived reason	NZ European/ Pākehā	Māori	Pacific	Asian	Other
Race	12%	33%	31%	27%	44%
Ethnicity	3%	36%	24%	42%	42%
Gender	9%	19%	13%	5%	-
Nationality	7%	21%	24%	22%	27%
Sexual orientation	10%	19%	13%	16%	-
Religion	26%	7%	26%	28%	15%
Age	6%	23%	15%	-	-
Disability	8%	-	10%	-	-
Political views	22%	15%	4%	23%	26%
Appearance	20%	23%	18%	21%	21%
Other	21%	10%	5%	-	-
Don't know	8%	10%	5%	14%	-

Base: Respondents who had experienced online hate speech in the last 12 months (112). Note multiple responses were allowed.

Reasons for being the target of online hate also differed according to age. In this respect, for young participants, those aged 18-29 years, sexual orientation was the most common reason followed by appearance and political views. In contrast, an older age group (30-39 years old) rated political views, religion and appearance as frequent reasons, whilst 40-to-49-year old respondents indicated that their ethnicity was the most common cause of being targeted with hateful digital communications followed by their religion and appearance.

On the other hand, participants who identified themselves as non-heterosexual indicated that their sexual orientation was the most common reason for receiving online hate speech. Meanwhile, for heterosexual respondents the main perceived reason for being targeted by hate speech was for religious reasons.

netsafe

Self-reported impact of online hate speech

Highlights
- More than half reported that online hate speech had a negative impact, mainly emotionally but also on their behaviour and daily activities.
- Descriptions of emotional impact included anger, sadness, fear, and frustration among others; however for some online hate also affected their social interactions, sleep and/or work.
- Despite being targeted, a third indicated not being personally affected by online hate.

The online survey also gathered qualitative data through an open-ended question. We asked those participants who said they were personally targeted with online hate (n=112) to describe the impact that this experience had on them. The purpose of the question was to collect data based on participants' own words and explore in more detail their experiences of online hate in relation to potential harm and distress.

Participants' comments were coded and then grouped into main categories and themes. Our analysis was reflective and iterative. We categorised participants' responses in terms of affective/emotional and behavioural impact. The following sub-sections describe these two aspects in more detail.

Of the 112 participants who responded to the question, 6 in 10 commented that online hate speech had either an affective and/or behavioural impact on them. A third of respondents, in contrast, mentioned that although they were targeted, they were not affected by the hateful digital communication – see Figure 7.

netsafe
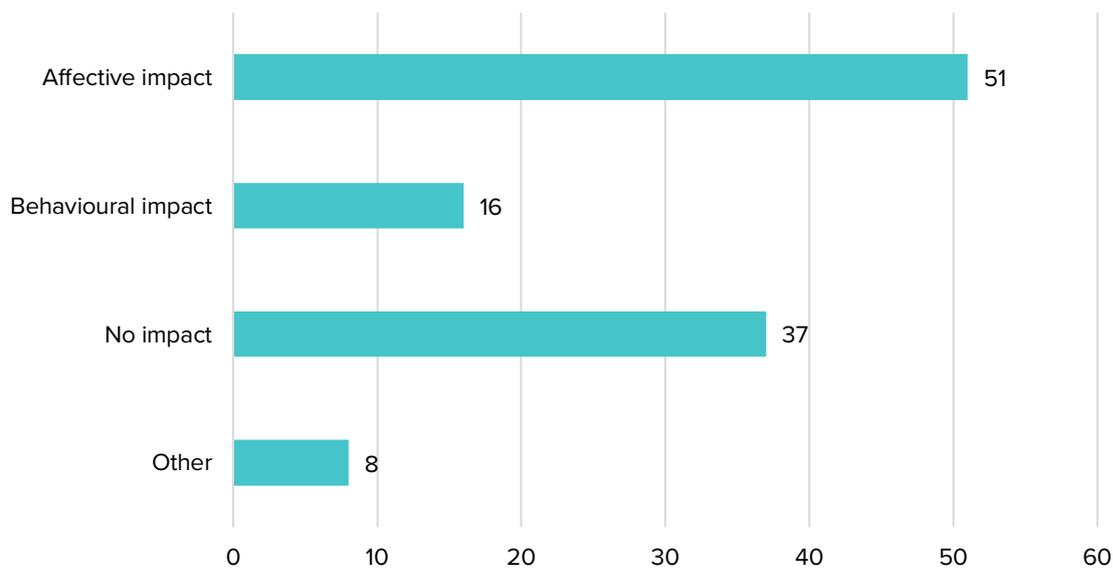
Figure 7. Number of participants self-reported impact of online hate speech

Base: Respondents who had experienced online hate speech in the last 12 months (112).

Affective impact. Of those who said they were negatively affected by online hate speech, most reported a range of emotional reactions. Those in this category described, for instance, having been "angry", "gutted", "upset", "stressed" or "anxious". A young Māori female respondent, who identified herself as lesbian, remarked, for example, that it was hurtful "that other people felt negative towards something I couldn't change about myself" and that it was worse when "such attitude was communicated online". Some participants who had moved to New Zealand also described the emotional impact caused by online hate. A young male (18-29 years old) from India, for example, commented that after being targeted with hateful content because of his nationality he felt "not welcome", and thus "not included as part of the wider community" despite "working hard" for it. For his part, another respondent, also from India, said he was "ashamed of being in New Zealand". A related experience was described by a New Zealand born female participant whose Cook Island Māori background, she said, was the reason for receiving online hate:

> When I first received these texts I was a bit confused…As a 59 year old in all my life I had never been told I was a fat ugly black coconut bitch that should go back where I came from. Hence the confusion. Everyone who knows me knows I am NZ born. By the 3rd text I had been called every bad name under the sun…when you are telling someone you are wrong several times you get really frustrated.

There were other participants who used terms such as "sad", "low", "down", "scared" or "humiliated" to describe the emotional impact of online hate speech. Among them, a young female participant indicated that she also felt "guilty", while another in the same age range (18-29 years old) was "very insecure". Both respondents commented they were targeted because of their appearance. Another participant reflected on how distressing online hate can be.

This middle-aged woman who identified herself as NZ European/Pākehā pointed out the following:

> *Sad, because you know these are people's true thoughts and opinions. People are faceless when they make comments on the internet, some shoot straight from the mouth, unaware of their broad (and silent) audience, others do it because they 'know' the audience is broad and that they are likely to attract attention.*

One aspect that surfaced from the data was not only the descriptions of emotional impact but also, in a few cases, participants' signs of resilience to deal with online hate. In this respect, one respondent said that he "was hurt inside" but then he "brushed it off". Another commented that although online hate "pissed me off", he "refused to dwell on it". A young female participant whose religion was the cause of receiving hateful content also said, "it made me feel a little sick and attacked" but soon she stopped paying attention to the situation.

Behavioural impact. For some participants the impact of online hate also extended to their everyday lives, in other words their behaviour and the activities they normally do. This was the case for a few participants who indicated avoiding "leaving the house sometimes" or being "scared to come out in public". A young female respondent said that she was "pushed away from family" while another commented that her sleep was not only affected, but also she started overeating to compensate for the distress. For some participants online hate speech affected work. Among them, a male (30-39 years old) who received hateful content targeting his nationality, commented that the experience "made me incredibly uncomfortable at work". For another, a middle-aged male respondent (40-49 years old), the impact was more severe as he stopped working after being targeted because of his disability.

For some there was a sense that online hate speech affected the way they interact with others and participate online. For example, a mature female participant (40-49 years old) said that "for a few months" she did not "talk to many people as I couldn't tell my side of things". Another female commented that online hate targeting her gender and age made her withdraw from the online group she used to participate in, and since then, she has "never returned, even though all the other people in that group were my friends". A third participant indicated she became "wary of what I share on social media".

Similarly, in response to receiving online hate others changed their daily use of digital technologies. This was the case for one respondent, a middle-aged Samoan participant, who said he now spends less time online and on the phone. Another participant commented that he was "angry" with online hate that targeted his political views and as a result he had "to stop using the website or app". Despite reporting "very little" emotional impact, one participant, on the other hand, indicated that he decided to unfriend from social media the person who sent him content targeting his political views and religion. Others preferred to block the sender. In another case, the positive role of online peer support was also highlighted over blocking or unfriending.

netsafe

This was the case for a middle-aged female respondent who was sent online hate because of her religion:

> I was somewhat distressed, but pleased to note that I did not have to reply, as others already intervened on my behalf.

On the other hand, another interesting aspect that emerged from peoples' comments was that, for some, online hate speech is part of the risks of interacting in the digital environment. One participant indicated that online hate is frustrating, but he had to accept it "as the norm for online interaction". Another participant complemented this point by saying that "it's the way life is now". A third respondent commented:

> I don't let it affect me – there is lots of opinions, rights and wrongs and a whole lot of people that enjoy others' misfortunes. I remind myself of that and continue to scroll.

Exposure to online hate speech

Highlights
- About 3 in 10 have seen or been exposed to online hate speech targeting someone else
- Exposure to online hate was more common among participants within 'other' ethnic groups

In addition to measuring personal experiences, the survey included a question intended to gauge exposure to online hate speech. More precisely, we asked participants to indicate whether they have seen or been exposed to a hateful digital communication that targeted someone or a group because of their race, ethnicity, gender, nationality, sexual orientation, religion, age, disability, and/or other attribute. All participants answered the question. While over a half of respondents (62%) indicated not having encountered this sort of online content in the last 12 months, 28% of them responded affirmatively. The remaining 10% of participants indicated they did not know whether they have been exposed to online hate – see Figure 8.

netsafe

YES
5% Once
14% A few times (2-4)
8% Many times (5+)
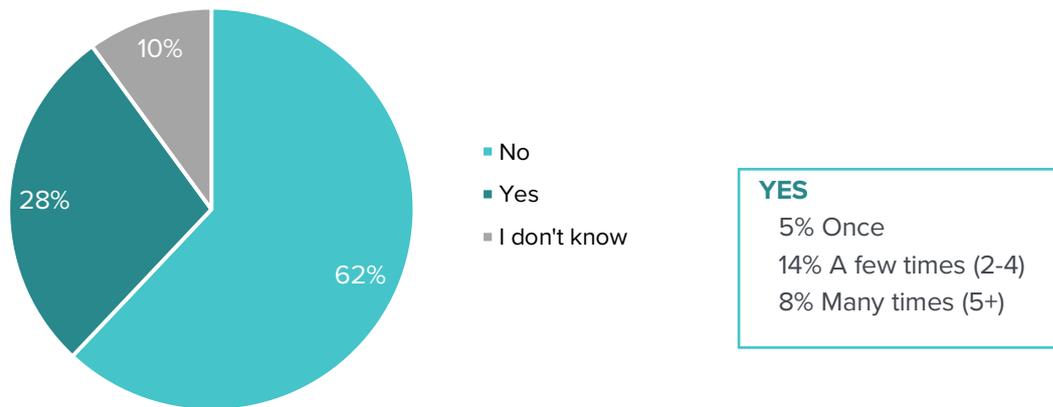
Figure 8. Overall exposure to online hate speech

Base: All respondents (1001).

Exposure to online hate speech was also more common among younger age groups. As Table 5 shows, those aged 18 to 29 (40%) and 30 to 39 years old (39%) reported the highest percentage of online hate exposure targeting others. These age groups were followed by those aged 40-49 years with 29% – see Table 5.

Table 5. Exposure to online hate speech by age

Frequency	18 – 29 years old	30 – 39 years old	40 – 49 years old	50 – 59 years old	60 – 69 years old	70 years or older
Never	46%	47%	59%	73%	80%	78%
Once	7%	7%	5%	3%	2%	4%
A few times (2-4)	19%	22%	16%	9%	6%	11%
Many times (5 or more)	14%	10%	8%	8%	4%	4%
I don't know	14%	13%	12%	8%	6%	4%

Base: All respondents (1001)

In terms of ethnicity, for participants who chose the category 'other' to describe their ethnicity, exposure to online hate speech was significantly higher compared to other ethnic groups (see Table 6). Some 41% of respondents in this group indicated having encountered or seen online hate one or more times in the past year. In comparison, percentages were lower but still significant among Māori (30%) and Asian (29%) followed by NZ European/Pākehā (27%) and Pacific (25%) participants.

netsafe

Table 6. Exposure to online hate speech by ethnicity

Frequency	NZ European/ Pākehā	Māori	Pacific	Asian	Other
Never	66%	53%	57%	53%	46%
Once	5%	4%	7%	7%	-
A few times (2-4)	14%	16%	12%	17%	22%
Many times (5 or more)	8%	10%	6%	5%	19%
I don't know	7%	17%	19%	18%	13%

Base: All respondents (1001)

On the other hand, the difference in online hate exposure in the context of gender was not statistically significant. In this respect, 29% of female participants and 27% of males indicated having encountered hateful content online in the past year. Regarding sexual orientation, over a half of non-heterosexual participants (55%) indicated having seen or encountered hateful content while 30% said they were not exposed to it and 15% indicated they did not know whether they have seen any. In the case of heterosexual respondents, exposure to online hate was less common (26%), with most participants in this group (65%) answering they have not seen it in the prior year.

Perceived reasons for online hate speech against others

Highlights
- Participants who were exposed to online hate believed that others were more commonly targeted because of their religion, race and/or ethnicity.
- For both NZ European/Pākehā and those who identified within the 'other' ethnic category the main reason was religion. Māori and Pacific respondents pointed out race whilst Asians highlighted ethnicity.

Survey participants who reported to have been exposed to online hate in the prior year (n=285) were asked to indicate the reasons they think others were targeted. Figure 9 describes that for these participants religion, race, and ethnicity were the most common motives behind the hateful content they were exposed to.

On the other hand, while different age groups mostly highlighted the same reasons behind online hate, the importance of these reasons was, in some cases, rated differently. For example, participants aged 18-29 years old indicated that the online hate they were exposed to targeted others because of their race, ethnicity, and sexual orientation. Meanwhile, those aged 30-39 and 40-49 years old considered that religion, race, and ethnicity were the main motives. For older participants (50-59 and 60-69) hateful content online was mainly motivated

netsafe

by religion, race and political views. Finally, participants over 70 years old mentioned religion, political views, and race.
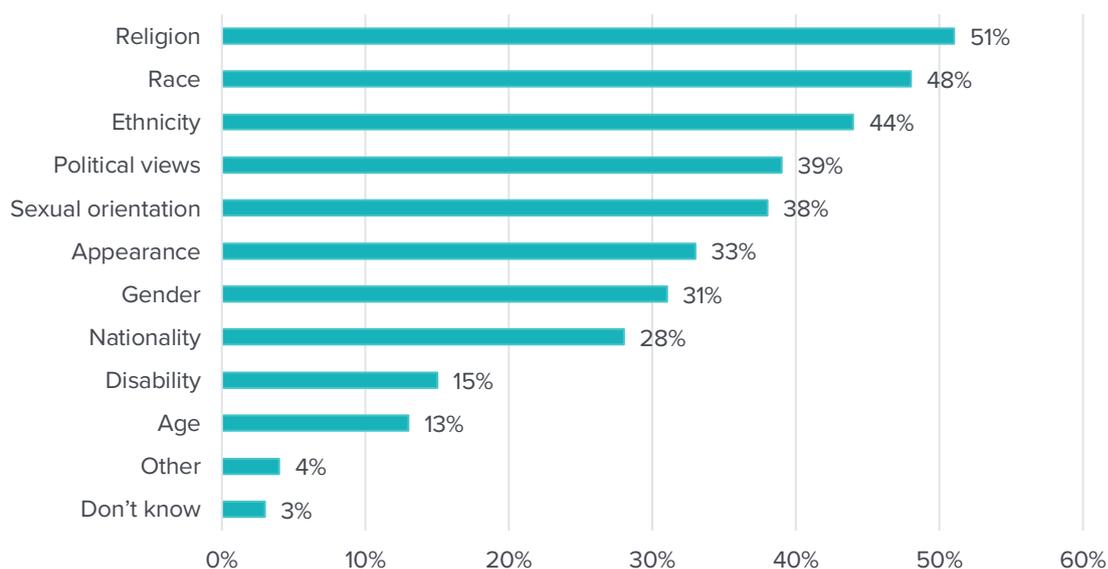
Figure 9. Overall perceived reasons for online hate speech against others

Base: Respondents who have seen online hate speech against others in the last 12 months (285). Note multiple responses were allowed.

In terms of gender, our findings indicate that, in general, male and female participants concurred on the reasons why others were the target of online hate speech (e.g. religion, race). However, interestingly, females' responses in relation to physical appearance were largely higher compared to males' answers. On the other hand, males' views about sexual orientation and gender as motivations for hateful online content against others were significantly higher compared to those from females. Further details are shown in Figure 10.
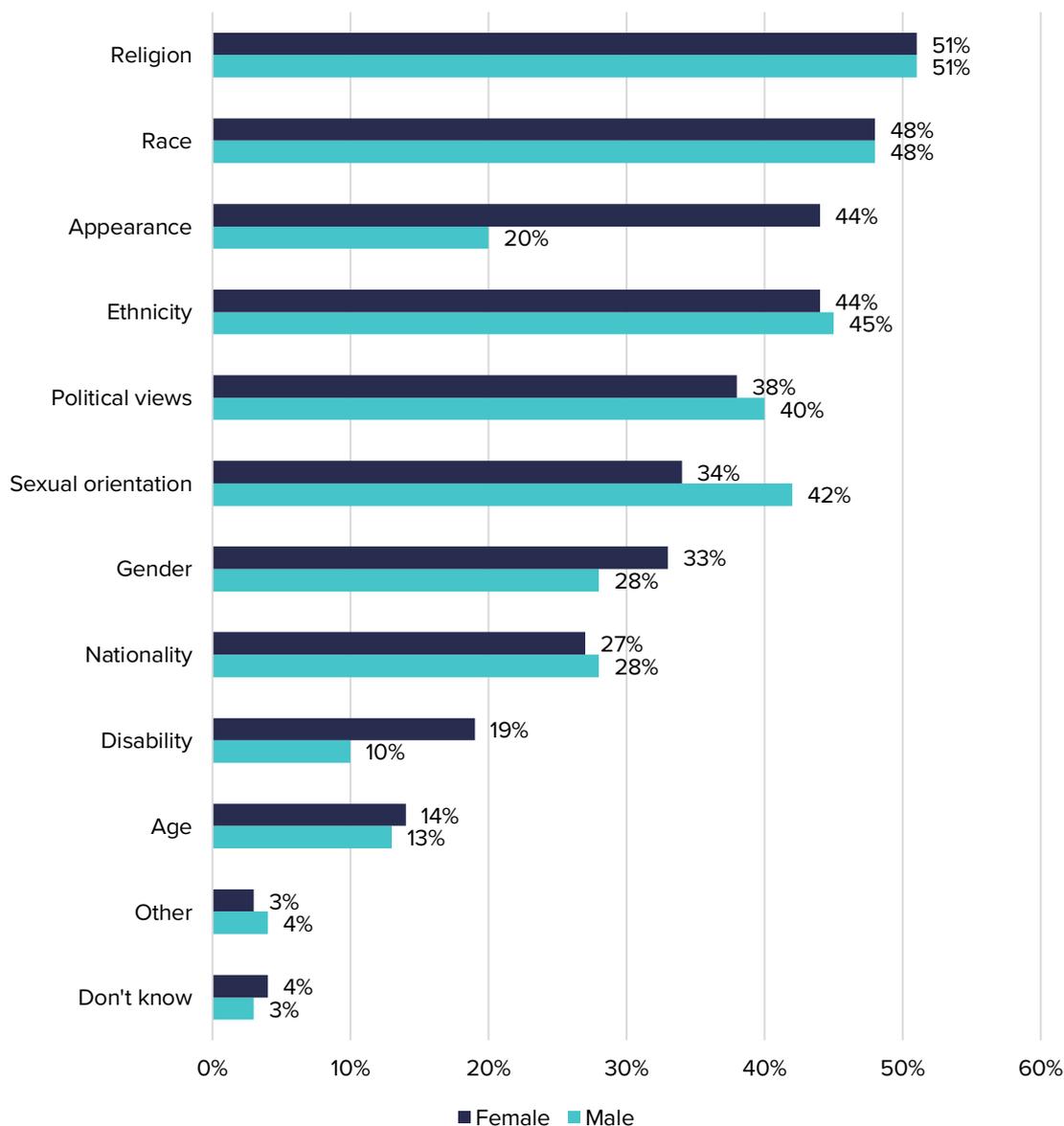
Figure 10. Perceived reasons for online hate speech against others by gender

Base: Respondents who have seen online hate speech against others in the last 12 months (285). Note multiple responses were allowed.

In terms of ethnicity, NZ European/Pākehā considered that the motivations behind the online hate they were exposed to followed this order: religion, race and political views. Māori, on the other hand, indicated that it was mainly race but also ethnicity and political views. Pacific respondents, meanwhile, pointed out race and then ethnicity along with sexual orientation and religion. For Asian participants who encountered online hate the reasons were ethnicity, race and religion. Finally, those who identified themselves as in an 'other' ethnic group highlighted the most common reasons for online hate in the following order: religion, race, and ethnicity. Details are given in Table 7.

Table 7. Perceived reasons for online hate speech against others by ethnicity

Ethnicity	NZ European/ Pākehā	Māori	Pacific	Asian	Other
Race	45%	59%	51%	46%	69%
Ethnicity	39%	50%	43%	56%	66%
Gender	32%	37%	15%	31%	39%
Nationality	26%	25%	25%	34%	35%
Sexual orientation	39%	43%	42%	28%	36%
Religion	50%	45%	42%	40%	76%
Age	13%	17%	9%	21%	16%
Disability	17%	21%	9%	9%	5%
Political views	43%	49%	22%	19%	41%
Appearance	34%	45%	21%	15%	44%
Other	5%	1%	5%	-	-
Don't know	4%	2%	2%	4%	-

Base: Respondents who have seen online hate speech against others in the last 12 months (285). Note multiple responses were allowed.

Engagement with online sites promoting/distributing online hate speech

Highlights
- In the prior year, 1 in 20 have intentionally visited a website, online forum and/or social media site that promotes or distributes online hate speech
- Among those who did so, the behaviour was more common among males, young people under 30 years old, and those who identified themselves as non-heterosexual

The last survey question asked participants whether in the prior year they have 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, intentional engagement with this type of online environment was uncommon among survey participants.

netsafe

As the results in Figure 11 show, only 5% of participants have deliberately visited online sites that promote hate speech against others.

YES
1% Once
2% A few times (2-4)
1% Many times (5+)
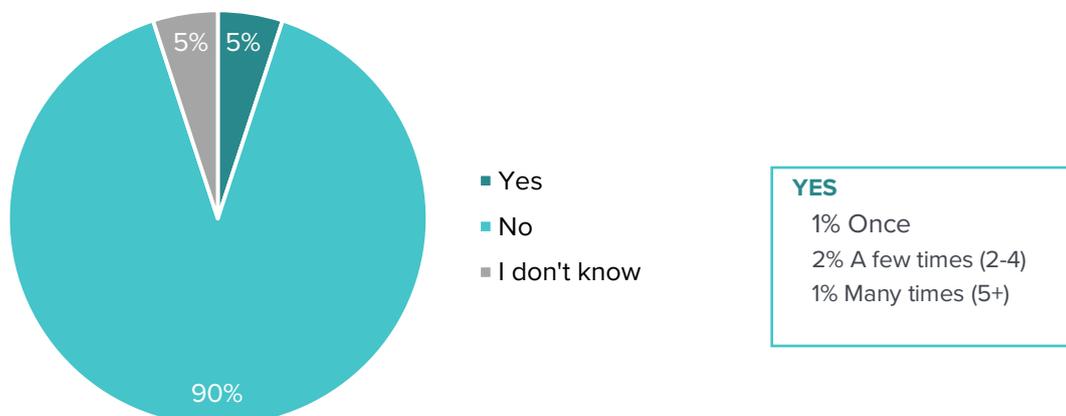
Figure 11. Overall engagement with online sites promoting online hate speech

Base: All respondents (1001).

Intentional engagement with online environments that promote hate speech varied depending on respondents' gender. Our data show that the proportion of males (6%) who engaged in this behaviour seemed to be double that of females (3%) – see Figure 12.
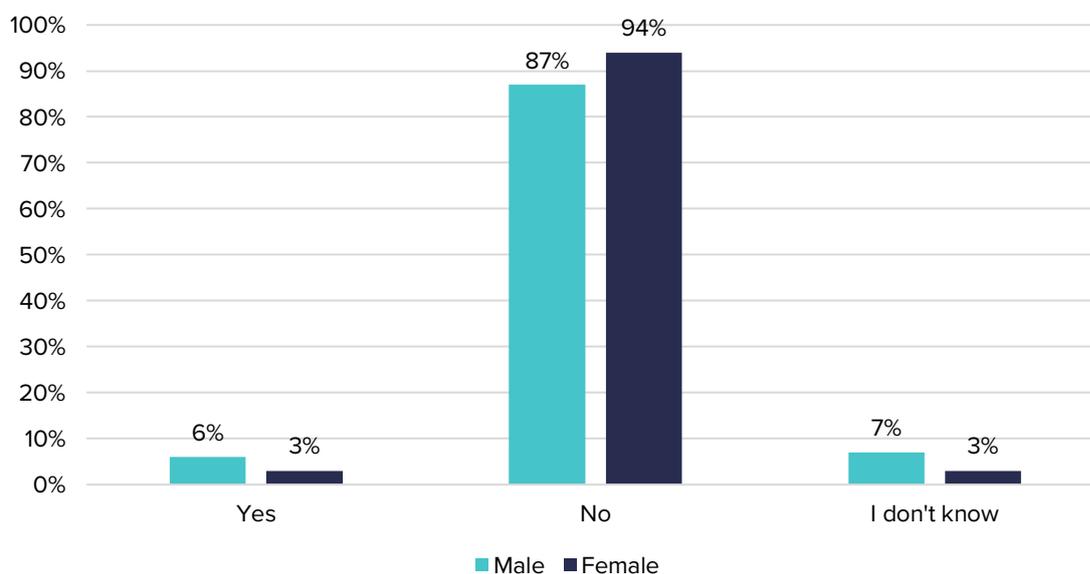
**Figure 12. Engagement with online sites promoting online hate speech by gender**

Base: All respondents (1001).

netsafe

In terms of age, our data show that engagement with online hate sites was higher among those aged under 30 years old. Some 8% in this group indicated that in the prior year they have intentionally visited an online space promoting hate speech one or more times – more details are given in Table 8.

Table 8. Engagement with online sites promoting online hate speech by age

Frequency	18 – 29 years old	30 – 39 years old	40 – 49 years old	50 – 59 years old	60 – 69 years old	70 years or older
Never	80%	87%	94%	93%	94%	97%
Once	2%	2%	1%	2%	-	2%
A few times (2-4)	5%	1%	1%	3%	2%	*
Many times (5 or more)	1%	2%	-	1%	-	-
I don't know	12%	8%	4%	1%	4%	-

Base: All respondents (1001).

*Non-statistically relevant number of responses

Table 9, on the other hand, shows engagement with online hate sites in terms of ethnicity. Overall, 10% of Asian participants indicated that they visited this sort of website one or more times in the prior year. This ethnic group was followed by Pacific respondents (7%). Note that statistical variances among some ethnic groups are modest.

Table 9. Engagement with online sites promoting online hate speech by ethnicity

Frequency	NZ European/ Pākehā	Maori	Pacific	Asian	Other
Never	93%	87%	87%	79%	87%
Once	1%	2%	2%	3%	1%
A few times (2-4)	1%	2%	4%	7%	4%
Many times (5 or more)	1%	1%	1%	-	-
I don't know	4%	7%	7%	11%	8%

Base: All respondents (1001).

netsafe

Discussion and Conclusion

This report has presented findings from an exploratory study about online hate speech in New Zealand. It has found that 1 in 10 adult New Zealanders report being personally targeted by online hate speech, and that 3 in 10 have seen or encountered hateful content online one or more times in the prior year. In addition to measuring prevalence and frequency, the results include insights about people's perceived reasons for being targeted, the impact of online hate as well as the intentionality of people's engagement with online spaces (e.g. forums, websites, social media) promoting hateful digital communications. Findings presented in this report are based on nationally representative data focusing on key demographics such as age, gender, and ethnicity.

A significant finding is that online hate speech in New Zealand is more likely to target minorities. Our research has uncovered that members of minority ethnic groups, particularly Asians, non-heterosexual people (e.g. gay, lesbian, bisexual), and people with disabilities were more commonly targeted with hateful content online. These findings relate to international research and debate about the prevalence of online hate against vulnerable groups (see Foxman & Wolf, 2013; Waldron, 2012). Equally important are the results that depict the negative impact of online hate. In most cases and to different degrees, adult New Zealanders described the emotional impact they have experienced as a result of online hate directed to them. Furthermore, as our qualitative data show, some have been more affected, resulting in their inability to socially interact or carry on with their everyday lives. These findings are an invitation for further reflection about online hate speech. While discussion in New Zealand has mainly looked at the legal and ethical challenges surrounding online hate, the social aspect that involves understanding the experiences, needs and views of specific groups, communities, and people as individuals also must be considered.

Similarly, the study has found that online hate varies with age. As the finding shows, personal experiences of and exposure to hateful content online are more common among youth and younger adults. This is consistent with international research (Costello et al., 2016; Hawdon et al., 2014; Kaakinen et al., 2018) and supports the idea that studying online hate in the context of young people is particularly relevant. An explanation for this finding might be that online hate groups actively seek to recruit or influence people from younger age groups, as some commentators suggest (see Blaya, 2018; Lee & Leets, 2002) and/or that youths tend to be actively engaged with digital tools, especially social media, compared with older age groups (Hawdon et al., 2014). This age difference regarding online hate also has implications for practice. Not only does online hate target – or is more often encountered by – young people but it is more likely to affect them negatively. A recent Netsafe study suggests, for instance, that young people in New Zealand are more vulnerable to the impact of unwanted digital communications (see Netsafe, 2018b).

Interestingly, a surprising finding relates to personal experiences of online hate and gender. Available research has shown that females are more likely to be on the receiving end of online harassment and other forms of risky and abusive online behaviours (Duggan, 2017; Henry & Powell, 2018; Ministry for Women, 2017; Netsafe, 2018b; Staude-Müller, Hansen, & Voss, 2012). We were expecting a similar pattern. However, our data show that, in the context of online hate speech, males were more commonly targeted than females. Although the gender difference in our study is relatively modest, this finding might be explained by the higher rate of males' responses about ethnicity, race and political views, which along with

netsafe

religion are top reasons for being targeted with online hate speech. However, on the other hand, it also must be noted that gender and age-related online hate were significantly more common among females.

Our study also offers interesting insights about who in New Zealand has intentionally visited an online site(s) that promotes or distributes hateful content. We know now that, in the last 12 months, about 5% of adults visited a website, online forum and/or social media site that promoted hateful content, and that this behaviour was more common among males, youths and young adults under 30 years old, Asians, and those who identified themselves as non-heterosexual. However, while these demographic characteristics are relevant, we are unable to make further interpretations and conclusions based on our data. We recommend interpreting the meaning of these figures cautiously. For example, intentionally visiting an online hate forum does not explain motivations for visiting a forum, level of engagement with its content and interaction with its members, nor alignment with the ideology promoted. The use of other dependent variables such as types of online engagement and political views will be required in further research to understand this matter in more depth.

Finally, as online hate speech is a highly complex topic, we acknowledge that our study only offers a general look at it and has some challenges and limitations, explained in the Methodology section of this report. We think the understanding of online hate demands the use of other types of methodologies and research techniques as well as rigorous exploration of other aspects of this phenomenon. For this reason, we suggest avenues for future research such as studying in more detail those groups more commonly targeted, the motivations to engage with this online behaviour as well as the effectiveness of approaches (computational, legal, educational) to prevent or mitigate online hate. Longitudinal research or research conducted during a specific timeframe (e.g. general elections) will also help to further understand the nature and impact of online hate speech in New Zealand.

netsafe

Glossary

Digital communication: A digital communication refers to any form of electronic communication as defined in the Harmful Digital Communications Act 2015. This includes any text message, writing, picture, recording, or other matter that is communicated electronically.

Digital harm and distress: To describe experiences of digital harm and/or distress we consider two elements: affective and behavioural. Affective is how a person describes feeling about a situation of unwanted digital communication (e.g. annoyed, anxious, unsafe). It is the internal experience of emotional reaction to it. Behavioural refers to the impact of an unwanted digital communication(s) on the person's daily activities (e.g. unable to leave the house as a result of being anxious or feeling unsafe).

Unwanted digital communications: Unwanted digital communications include a range of online experience(s) mediated/facilitated by unsolicited electronic communication(s) that might or might not cause distress and/or harm to the person who deals with it (e.g. receiving spam, accidentally seeing inappropriate content, having rumours spread about oneself, being threatened online).

Gender: The social and cultural construction based on the expectation of what it means to be a man and/or a woman, including roles, expectations, and behaviour. The concept of gender diversity acknowledges this full range of genders. Societies, and cultures within societies, have different constructs and expectations of gender and this can vary over time (Statistics NZ, n.d.).

Sexual orientation: A person's sexual orientation can be derived from their sexual attraction, sexual behaviour, and/or sexual identity. It includes, for example, lesbian, gay, bisexual, queer, heterosexual, and asexual (Statistics NZ, n.d.).

netsafe

Acknowledgements

Our gratitude to Distinguished Professor Paul Spoonley, Pro Vice-Chancellor at Massey University, Dr. Philippa Smith, Senior Lecturer at Auckland University of Technology, Taine Polkinghorne, Human Rights Advisor from the Human Rights Commission, and Angela Webster (MA) who is Netsafe's Education Advisor, for their valuable feedback at different stages of this study. The views and interpretations expressed in this report are those of Netsafe.

Further Resources

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 or visiting https://netsafe.org.nz/report

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/

To contact Netsafe for more information about its research programme email research@netsafe.org.nz

netsafe

References

Alba, D. (2017). Defining "hate speech" online is an imperfect art. Retrieved from
https://www.wired.com/story/defining-hate-speech-online-is-imperfect-art-as-much-as-science/

Blaya, C. (2018). Cyberhate: A review and content analysis of intervention strategies. *Aggression and Violent Behavior*. https://doi.org/10.1016/J.AVB.2018.05.006

Brown, A. (2017). What is so special about online (as compared to offline) hate speech? *Ethnicities*, 146879681770984. https://doi.org/10.1177/1468796817709846

Bryman, A. (2008). *Social research methods* (Third edition). New York: Oxford University Press.

Buranyi, S. (2017). Rise of the racist robots – how AI is learning all our worst impulses. Retrieved October 17, 2018, from https://www.theguardian.com/inequality/2017/aug/08/rise-of-the-racist-robots-how-ai-is-learning-all-our-worst-impulses

Burris, V., Smith, E., & Strahm, A. (2000). White supremacist networks on the internet. *Sociological Focus*, *33*(2), 215–235. https://doi.org/10.1080/00380237.2000.10571166

Costello, M., Hawdon, J., Ratliff, T., & Grantham, T. (2016). Who views online extremism? Individual attributes leading to exposure. *Computers in Human Behavior*, *63*, 311–320. https://doi.org/10.1016/J.CHB.2016.05.033

Costello, M., Hawdon, J., & Ratliff, T. N. (2017). Confronting online extremism: The effect of self-help, collective efficacy, and guardianship on being a target for hate speech. *Social Science Computer Review*, *35*(5), 587–605. https://doi.org/10.1177/0894439316666272

Duffy, M. E. (2003). Web of hate: A fantasy theme analysis of the rhetorical vision of hate groups online. *Journal of Communication Inquiry*, *27*(3), 291–312. https://doi.org/10.1177/0196859903252850

Duggan, M. (2017). Online harassment 2017. Retrieved August 28, 2017, from http://www.pewinternet.org/2017/07/11/online-harassment-2017/

eSafety Commissioner. (2017). Young people's experiences with online hate, bullying and violence. Retrieved October 16, 2018, from https://www.esafety.gov.au/education-resources/iparent/online-hate-infographic

Evans, J. R., & Mathur, A. (2005). The value of online surveys. *Internet Research*, *15*(2), 195–219. https://doi.org/10.1108/10662240510590360

Facebook. (n.d.). Hate speech – community standards. Retrieved October 5, 2018, from https://www.facebook.com/communitystandards/hate_speech

Foxman, A. H., & Wolf, C. (2013). *Viral hate: Containing its spread on the internet*. New York: Palgrave Macmillan.

Gagliardone, I., Gal, D., Alves, T., & Martinez, G. (2015). Countering online hate speech. Retrieved April 16, 2018, from http://unesdoc.unesco.org/images/0023/002332/233231e.pdf

Gerstenfeld, P. B., Grant, D. R., & Chiang, C.-P. (2003). Hate online: A content analysis of extremist internet sites. *Analyses of Social Issues and Public Policy*, *3*(1), 29–44. Retrieved from https://pdfs.semanticscholar.org/f14e/b2a9af434328ea458b8d6aedba2979aede12.pdf

HateBase. (2018). Most common hate speech. Retrieved September 26, 2018, from https://www.hatebase.org/popular

netsafe

Hawdon, J., Oksanen, A., & Räsänen, P. (2014). Victims of online hate groups: American youth's exposure to online hate speech. In J. Hawdon, J. Ryan, & M. Lucht (Eds.), *The causes and consequences of group violence: From bullies to terrorists* (pp. 165–182). Lanham, MD: Lexington Books.

Hawdon, J., Oksanen, A., & Räsänen, P. (2017). Exposure to online hate in four nations: A cross-national consideration. *Deviant Behavior*, *38*(3), 254–266. https://doi.org/10.1080/01639625.2016.1196985

Henry, N., & Powell, A. (2018). Technology-facilitated sexual violence: A literature review of empirical research. *Trauma, Violence, & Abuse*, *19*(2), 195–208. https://doi.org/10.1177/1524838016650189

Jones, L. M., Mitchell, K. J., & Finkelhor, D. (2013). Online harassment in context: Trends from three Youth Internet Safety Surveys (2000, 2005, 2010). *Psychology of Violence*, *3*(1), 53–69. https://doi.org/10.1037/a0030309

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

Keipi, T., Näsi, M., Oksanen, A., & Räsänen, P. (2016). *Online hate and harmful content: Cross-national perspectives*. London: Routledge. https://doi.org/10.4324/9781315628370

Lee, E., & Leets, L. (2002). Persuasive storytelling by hate groups online. *American Behavioral Scientist*, *45*(6), 927–957. https://doi.org/10.1177/0002764202045006003

Livingstone, S., Haddon, L., Görzig, A., & Ólafsson, K. (2011). Risks and safety on the internet: The perspective of European children: Full findings and policy implications from the EU Kids Online survey of 9-16 year olds and their parents in 25 countries. Retrieved September 5, 2017, from http://eprints.lse.ac.uk/33731/

Martin, F. R., & Rolph, D. (2016). Removing social media hate speech within 24 hours sounds like a good idea, but…. Retrieved from https://theconversation.com/removing-social-media-hate-speech-within-24-hours-sounds-like-a-good-idea-but-60325

Ministry for Women. (2017). *Insights into digital harm: The online lives of New Zealand girls and boys*. Wellington, New Zealand. Retrieved from http://women.govt.nz/documents/insights-digital-harm-online-lives-new-zealand-girls-and-boys

Miškolci, J., Kováčová, L., & Rigová, E. (2018). Countering hate speech on Facebook: The case of the Roma minority in Slovakia. *Social Science Computer Review*. https://doi.org/10.1177/0894439318791786

Mondal, M., Silva, L. A., & Benevenuto, F. (2017). A measurement study of hate speech in social media. https://doi.org/10.1145/3078714.3078723

Netsafe. (2018a). *Harmful digital communications in New Zealand: Annual Population Survey 2017*. Wellington, New Zealand. https://doi.org/10.2139/ssrn.3128121

Netsafe. (2018b). New Zealand teens and digital harm: Statistical insights into experiences, impact and response. Wellington, New Zealand. https://doi.org/10.2139/ssrn.3188608

Oksanen, A., Hawdon, J., Holkeri, E., Näsi, M., & Räsänen, P. (2014). Exposure to online hate among young social media users. In M. N. Warehime (Ed.), *Soul of society: A focus on the lives of children and youth* (pp. 253–273). Oklahoma City: Emerald.

netsafe

Schmidt, A., & Wiegand, M. (2017). A survey on hate speech detection using natural language processing. In *Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media* (pp. 1–10). Retrieved from http://www.aclweb.org/anthology/W17-1101

Silva, L., Mondal, M., Correa, D., Benevenuto, F., & Weber, I. (2016, March 24). Analyzing the targets of hate in online social media. Retrieved August 14, 2018, from http://arxiv.org/abs/1603.07709

Spoonley, P. (2018). Te Whakamauāhara ki te Ipurangi: Hate speech in the age of the internet. Palmerston North: Massey University.

Statistics NZ. (n.d.). Gender identity – Glossary and references. Retrieved from http://archive.stats.govt.nz/methods/classifications-and-standards/classification-related-stats-standards/gender-identity/glossary-references.aspx

Staude-Müller, F., Hansen, B., & Voss, M. (2012). How stressful is online victimization? Effects of victim's personality and properties of the incident. *European Journal of Developmental Psychology*, *9*(2), 260–274. https://doi.org/10.1080/17405629.2011.643170

Waldron, J. (2012). *The harm in hate speech*. Cambridge, Massachusetts: Harvard University Press.

Wansink, B. (2001). Editorial: The power of panels. *Journal of Database Marketing & Customer Strategy Management*, *8*(3), 190–194. https://doi.org/10.1057/palgrave.jdm.3240034

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

Ybarra, M. L., Mitchell, K. J., & Korchmaros, J. D. (2011). National trends in exposure to and experiences of violence on the Internet among children. *Pediatrics*, *128*(6), e1376–86. https://doi.org/10.1542/peds.2011-0118

YouTube. (n.d.). Hate speech policy – YouTube Help. Retrieved October 5, 2018, from https://support.google.com/youtube/answer/2801939?hl=en

netsafe

Appendix: Online hate speech questionnaire

Module E - Online hate speech *[extracted from the APS questionnaire]*

Now we have some questions about digital communications that offend, discriminate, denigrate, abuse and/or disparage people because of race, ethnicity, gender, nationality, sexual orientation, religion, age, disability, and/or others.

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)

Never	SKIP TO Q43
Once	
A few times (2-4)	
Many times (5 or more)	
I don't know	SKIP TO Q43

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.

[Note: ROTATE ORDER EXCEPT FIX BOTTOM 2].

Race
Ethnicity
Gender
Nationality
Sexual orientation
Religion
Age
Disability
Political views
Appearance
Other. Please explain _____
Don't know

Q42 - In your own words, can you describe the impact this/these digital communication(s) had on you?

_____

_____

netsafe

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?

Never   SKIP TO Q45

Once

A few times (2-4)

Many times (5 or more)

I don't know   SKIP TO Q45

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.

[Note: ROTATE ORDER EXCEPT FIX BOTTOM 2].

Race

Ethnicity

Gender

Nationality

Sexual orientation

Religion

Age

Disability

Political views

Appearance

Other. Please explain _____

Don't know

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?

Never

Once

A few times (2-4)

Many times (5 or more)

I don't know

netsafe