

# THE PEOPLE'S REPORT ON ONLINE HATE, HARASSMENT AND ABUSE



THE PEOPLE’S REPORT ON ONLINE  
HATE, HARASSMENT, AND ABUSE

---

|                        |    |
|------------------------|----|
| FOREWORD               | 4  |
| INTRODUCTION           | 6  |
| WHAT OUR FINDINGS SHOW | 10 |
| RECOMMENDATIONS        | 36 |
| NOTES ON METHODOLOGY   | 38 |
| ACKNOWLEDGEMENTS       | 46 |

*Netsafe offers a free service for people in New Zealand to help with online bullying, harassment and abuse – it is available seven days a week on **0508 NETSAFE** (058 638 723). If it is an emergency, please call **111**.*

# FOREWORD

## FOREWORD

Behaviour intolerable in real life is somehow easier online.

Hate, abuse and harassment happen too frequently online and with a fierce intensity, especially to marginalised people.

This report adds to the understanding of the issues with New Zealand data, and a direction to update our outdated rules and regulations.

Being online is a misnomer. It's like walking on footpaths and driving on roads -- part of everyday life. Yet we seem to treat online as a separate space rather than an extension of everyday life.

The anonymity easily found online can be a shield for hateful and abusive behaviour. Social norms that govern our data to day interactions do not always apply online.

Careful academic work around the world shows the impacts of harmful online behaviour are real. Startlingly, many young people find online bullying much more harmful than social bullying.

The online space has rapidly become an extension of our daily lives, but our rules, regulations and enforcement have not kept pace.

Old problems of antisocial behaviours are not easily addressed.

We need to think deliberately and creatively to deal with old problems in a new medium.



**Shamubeel Eaquib** is an experienced economist who makes economics easy. He is also an author, media commentator and a thought leading public speaker. With Netsafe he authored the report *Cyberbullying- Estimating Costs*.



# INTRODUCTION

## INTRODUCTION

For decades, the internet has been hailed as a groundbreaking interactive marketplace of ideas, where anyone with access to data and a device can set up a stall.

Online tools have made it possible to communicate easily with friends and whānau around the world, sell and purchase goods and services, enrol to vote, raise billions for charitable causes or start-up businesses, and even hail a ride or meal to your front door.

Right now, a child in Whangārei with a smartphone and WiFi has access to more information and better mobile communication than the US President did just 15 years ago.

The internet has helped give people who have historically been locked out of democracy by discrimination or poverty a way to voice the needs of their communities and organise at scale.

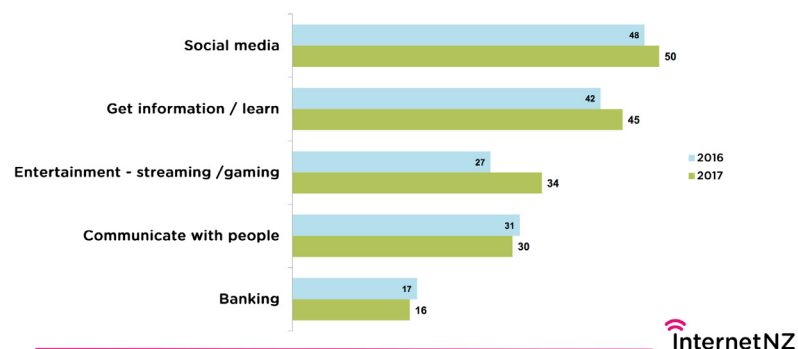
Over the past four years, ActionStation members have used digital tools and platforms to connect and collaborate with hundreds of thousands of other New Zealanders who share their vision and values to engage powerfully in our democracy.

Massive crowdsourced submissions to Select Committees have been drafted in Google Docs, and custom-built websites have enabled hundreds of people to contribute their stories to participatory research like the [\*People's Mental Health Report\*](#), the [\*People's Review of Renting\*](#) and the [\*They're Our Whānau\*](#) report. Using online surveys and discussion forums, around 10,000 New Zealanders contributed to a crowdsourced vision for the future of our country, [\*Te Ira Tāngata\*](#). All of this work is funded by donations made largely online, and the ActionStation community has grown to 240,000 as word has spread about these projects through social media and email.

## INTRODUCTION

Facebook says 3.5 million New Zealanders had an active account in 2018 with over 2.3 million people logging in everyday. For those aged 25 and under, the primary source of entertainment and news is online. 94% of New Zealanders check the Internet at least once a day with social media being the main use for most people.

### Top 5 Internet uses



In the 21st century, social media has become the new public square.

The downside to this unparalleled information exchange and connectedness is that the internet also provides a powerful and relatively cheap way for groups and individuals to spread hate, fear, abuse and mis/dis/mal-information across time and space, and without transparency.

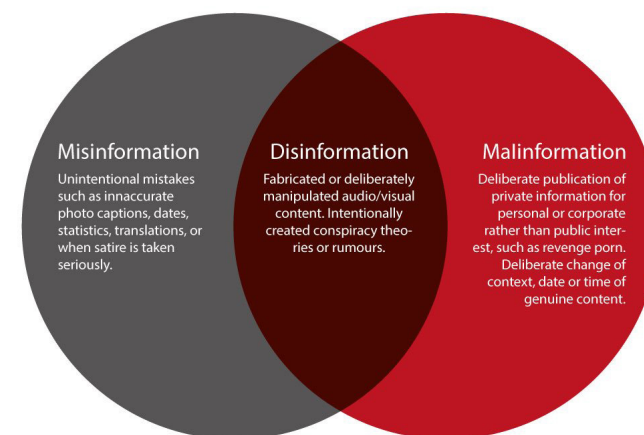
The term 'fake news' has been widely used to refer to a range of different kinds of false and harmful information. Clarity about those different types is important for policy making, so we will describe the differences between these three types of information using the following framework:

While ActionStation has been at the forefront of exploring and facilitating digitally-enhanced democratic participation in New Zealand, we have also been exposed to these downsides.

## TYPES OF INFORMATION DISORDER

FALSENESS

INTENT TO HARM



Source: *Information Disorder: Toward an interdisciplinary framework for research and policy making*

It is that exposure that has prompted this report.

In 2015, the National-led government passed the Harmful Digital Communications Act (HDCA). It states that a digital communication should not:

“...denigrate a person’s colour, race, ethnic or national origins, religion, gender, sexual orientation or disability.”

In 2019 we ask: has the Act worked? Is the internet free from prejudice and harm? Do people feel safe to participate freely in conversations online? Or is there more work to do?

This research seeks to better understand the impact of online hate, harassment and abuse on everyday New Zealanders. We put it into the world to add to the growing body of evidence demonstrating that the New Zealand government and the tech corporations, which now permeate so much of our lives, urgently need to do more to protect us online.

# WHAT OUR FINDINGS SHOW

## WHAT OUR FINDINGS SHOW

### ONLINE HATE IS COMMON AND WIDESPREAD, BUT IT'S WORSE FOR PEOPLE OF COLOUR, YOUNG PEOPLE, LGBT FOLK AND WOMEN

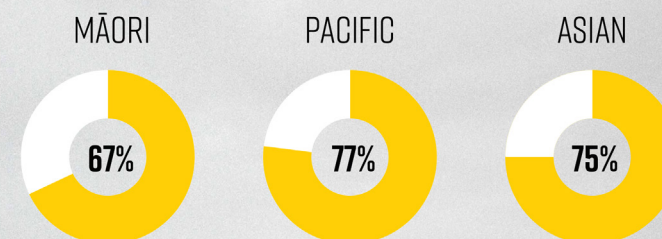
According to research from [Netsafe](#), one in ten New Zealanders experienced hate speech online and three in ten encountered hateful content in 2017.

LGBT people experience higher rates of online hate and abuse than others. And nearly two in ten (19%) New Zealand teenagers received an unwanted digital communication that had a negative impact on their daily life in 2018. It's worse for teenage girls, teenagers who are Māori or [teenagers with a disability](#).

Research by Amnesty International New Zealand found one in three New Zealand women experience online abuse and harassment. Of those women, three in four (75%) said they had not been able to sleep well, one in two (49%) feared for their physical safety and one in three (32%) feared for the physical safety of their families as a result.

[Our own research](#) shows that one in three Māori (32%), and one in five Asian (22%) and Pacific (21%) people experienced racial abuse and harassment online in 2018. The survey asked 618 Māori, Pacific and Asian people for their experiences, and the results were weighted by age, gender, and ethnicity.

#### Experienced racial abuse/harassment online



Base: All respondents (Māori n=210, Pacific n=205, Asian n=203)

## WHAT OUR FINDINGS SHOW

The survey asked about social media usage, experiences of online harassment and the impact of that harassment.

Of those people who had an experience of harassment most had faced racist abusive comments, messages or images directed at them.

[Social Media Mob: Being Indigenous Online](#) looks at the role social media plays in the lives of Indigenous Australians. Sadly, almost all participants had seen anti-black racism on Facebook and Twitter, making it clear that social media is not necessarily a safe space for Indigenous people in Australia. Further research into exposure to racist online hate may (and likely will) find that to be the case in New Zealand too.

## SOCIAL MEDIA PLATFORMS ARE EXACERBATING POLARISATION, AND IT COULD GET MUCH WORSE

2017 research conducted by Unitec into perceptions of community safety in West Auckland found that high social media use, particularly of Facebook and Neighbourly, increased Pākehā people's fear of crime, despite being the group least affected and despite crime rates consistently decreasing.

Pākehā people's use of social media was significantly higher than other ethnic groups (89% vs 69%) as was their anxiety and fear about other people in their community. Pākehā, as a group, wanted more police and tougher justice policy, compared with other ethnic groups who wanted more initiatives that support community engagement and connection.

A [report co-authored by former Chief Science Advisor Sir Peter Gluckman](#) concludes that:

*"Social media has helped new groups and communities of interest to form and grow, but it has also enabled more polarised and entrenched views to take root, and among the groups that form are antisocial ones that take advantage of the dark web to build new collectives based on violent, anti-social, morally objectionable or illegal behaviours."*

The report goes on to say:

*"Cyberbullying, and misuse of social media increasingly can polarise and fragment societies and threaten our sense of wellbeing, especially if they spill over into real-world actions. In the area of interpersonal relationships, the transition of online relationships to 'in real life' situations can have security consequences. It could be a minor who is enlisted into risk-taking behaviour in real life, or someone who has been radicalised online and goes on to violently enact their beliefs. In these and other types of cases, the persuasive and self-reinforcing nature of online activity carries an increased potential for real world consequences as the lines between our real and online lives increasingly blur."*

## ONLINE HARASSMENT TAKES MANY FORMS, AND THE IMPACTS ARE WIDE-RANGING

Between September and November 2018 we conducted an [online survey](#) to gather New Zealanders' stories of online harassment.

We asked 12 questions covering the context of the abuse, the platform, and the abuser. We had 69 in depth responses from around New Zealand.

Responses detailed a wide-range of forms of abuse: bullying and harassment from ex-partners, abuse that carried over from in person relationships, anonymous harassers, communities targeting individuals online and organised campaigns of harassment.

*They looked at my profile to check my home town then threatened to come "kick my teeth in" as they live 30 minutes away.*

*Some people have singled me out for abuse during some topics because I am male, some because I am Māori, some others because I am also NZ European and "sexual assault can't happen to a male".*

*It was a group of people and it was definitely organised. They told us as much. They had a forum of some kind where they'd decide who would be their victim and coordinate their attacks to do as much damage as possible.*

*On one occasion it was an organised attack, I got hundreds of twitter @s for tweeting about women's clothes often not having proper pockets being everyday sexism and got RTed by a hate account that specifically targets a few people who use the hashtag everyday, with their followers then sending abuse.*

The abuse takes place across the web in comments on blogs, in emails, group messages and in videos. Social media was mentioned as the platform where it took place most often, with Facebook mentioned more than any other platform. It's worth noting that Facebook is the [second most popular website in New Zealand after Google](#).

The negative impacts on people as a result of online abuse are also wide-ranging:

In our UMR survey results we found that one in four Māori (26%), one in three Asian (33%) and one in three Pacific (34%) people had lower self esteem and/or a loss of confidence as a result of online harassment.

Almost two in ten Māori (19%) and Pacific (18%) people, and one in three Asian (34%) reported a feeling of powerlessness in their ability to respond to the abuse or harassment.

Most who faced harassment changed their behaviour because of it. One in three Asian people (36%) used social media less as a result of the abuse. Almost one in three Pacific people (30%) made their accounts private while 14% of Māori stopped posting content that expressed their views. The abuse was reported to happen most often on Facebook, the most popular social network amongst those surveyed and in the country.

*I post less and less about personal things and wish to talk to or socialise with other people less. I often refuse to admit how I really am to people out of fear that it will be used against me for other peoples personal gain.*

## OUR LAWS ARE NOT ADEQUATELY PROTECTING VICTIMS OF DIGITAL DOMESTIC VIOLENCE

A number of the responses to our survey provided examples of ways in which offline abuse, particularly between intimate partners, extends to harassment and abuse online.

*It was my ex-partner so it was personal. But it was a concerted campaign.*

*A vindictive stalking ex partner smeared me and my children, swore at me and got others to join in*

*An ex-boyfriend hijacked my nieces account and posted content that was damaging.*

Lawyer Ruby King notes in her paper Digital Domestic Violence: Are victims of intimate partner cyber harassment sufficiently protected by New Zealand's current legislation?:

*"Levels of domestic violence in New Zealand are staggeringly high. Intimate partner cyber harassment represents a new form of domestic violence made possible by new technology and our increasing reliance on it. Further, as technology is being used in the formation and development of relationships and personal boundaries, there is a risk that intimate partner cyber harassment will become "normalised." Due to the complex and harmful nature of the issue, New Zealand's legislative regime must adequately protect victims."*

She also notes that:

*"The demographic most vulnerable to intimate partner cyber harassment is young females. While both intimate partner violence and cyber harassment are gender-neutral crimes, statistics clearly show that females are more likely to be victimised."*

## CYBERBULLYING IS COSTING THE COUNTRY MILLIONS, POTENTIALLY BILLIONS

[Netsafe and Sense Partners' research](#) conservatively estimates that cyberbullying costs New Zealanders \$444m a year. That figure includes the cost of interventions and loss of life, but does not include the effects of cyberbullying on long term health or productivity, so the actual figure is likely higher.

[Research conducted in Australia](#) found 39% of people have experienced one or more forms of online harassment, equivalent to 8.8 million Australians. Based on this research, the estimated total cost of cyber trolling to the national economy in terms of medical expenses and lost income sat at a staggering \$3.7 billion.

## MISINFORMATION SPREAD ONLINE CAN FUEL HATE AND DAMAGE DEMOCRACY

According to The Washington Post, 2018 was the year of online hate.

Between July and September 2018 Facebook removed nearly three million pieces of hate speech, while YouTube removed more than 140,000 videos featuring “hateful or abusive” content.

Over the last two decades, social media corporations have risen to prominence with a “move fast and break things” attitude, and while this is no longer Facebook’s company motto, this approach to building tech platforms was widely adopted and the results of that mentality remain.

During the 2016 US elections Russian disinformation groups used Facebook, Twitter, YouTube (owned by Google) and Instagram (owned by Facebook) to target specific groups with curated content meant to anger them and drive their voting behaviour.

According to Oxford University:

*“Russia’s Internet Research Agency (IRA) activities were designed to polarise the US public and interfere in elections by campaigning for African American voters to boycott elections or follow the wrong voting procedures in 2016... for Mexican American and Hispanic voters to distrust US institutions; encouraging extreme right-wing voters to be more confrontational; and spreading sensationalist, conspiratorial, and other forms of junk political news and misinformation to voters across the political spectrum.”*

Over 30 million people shared the IRA’s Facebook and Instagram posts with their friends and family between 2015 - 2017. This orchestrated manipulation of social media by Russian operatives, the majority of which revolved around dividing the nation along racial lines, was later connected to an increase in racialised hate crimes (Center for the Study of Hate and Extremism; California State University).

In 2017 a violent and ultimately deadly neo-Nazi march in Charlottesville, USA was planned, advertised, coordinated and paid for using PayPal, Facebook, and the gamer chat app Discord.

In 2018 misinformation spread through Facebook-owned messaging platform WhatsApp incited at least 27 lynchings in India. Meanwhile a United Nations human rights expert concluded Facebook helped facilitate “acts of genocide” in Myanmar. Cambridge Analytica harvested the personal data of millions of people’s Facebook profiles without their consent and used it for political purposes. A subsequent attack gave unknown hackers full access to 50 million people’s accounts, including private messages, location information and the ability to make posts.

These dangerous, and in some cases deadly, outcomes of the tech behemoths “move fast and break things” approach should be enough to force governments everywhere to take urgent action. Just as the internet has created immense positive value, it has also given new tools to people and groups who want to hack, silence, threaten, harass, intimidate, defame, or violently attack other people.

While there is limited research into the effects and reach of online mis/dis/mal-information in New Zealand, we are not immune to the threats that we’ve seen play out overseas.

Far-right fringe hate groups (largely based in the US) targeted French and German elections to influence the outcome. The EU commission has told US tech giants Facebook, Google, Twitter, Mozilla and advertising businesses to intensify their actions against disinformation campaigns ahead of European elections in May 2019, or face regulation.

The dangerous use of ‘dark ads’ (online ads which can only be seen by the people they are targeting) by bad faith actors has been well documented overseas. We have no system to document and oversee these ads in New Zealand. In the 2017 general election political parties spent hundreds of thousands of dollars on Facebook advertising. There is no record of the content of those ads or who they were targeted at. There are also few, if any, restrictions or disclosure requirements on the type of audience being targeted in ads.

Coordinated anti-1080 campaigns facilitated by Facebook, often based on misinformation, are an example of the kind of momentum created when groups and individuals can find widespread support online.

We have also seen a troubling rise in the toxicity of online discussions. Politicians, journalists and public figures who appear in the media, as well as anyone who shares their views on current affairs, often become the focus of targeted online harassment and abuse campaigns; it’s especially bad for women.

This abuse could discourage people from taking part in the conversations that help shape public opinion and influence decisions, posing a real threat to the functioning of our democracy.

Giant tech corporations like Facebook, Google and Twitter must face the public and political scrutiny worthy of infrastructure which billions of people now rely on for employment, information and social interaction.

As noted in the InternetNZ discussion paper *Platforms and misinformation: Where we are and how we got here:*

*“The Internet did not create disinformation. It did not create divisions between groups of people who are different, or have different ideologies about the world. It did, however, create new ways to spread this information, and much faster than we have ever seen before.”*

This needs to be an area of focus for the New Zealand government.

## YOUTUBE IS FACILITATING THE SPREAD OF MISINFORMATION

In July 2018 New Zealand was hit by a storm of debate about ‘hate speech vs free speech’ created by Lauren Southern and Stefan Molyneux, two controversial YouTubers known best for their anti-Muslim, anti-women and anti-multiculturalism views. The pair had planned a visit to Auckland for a speaking event.

The controversy arose after Auckland Council denied them a council venue to speak at and members of the public organised a well-attended peaceful protest against their visit.

But for anyone paying attention to the online movements that have grown in the wake of the toxic Gamergate campaign, the pair’s views and actions were likely unsurprising.

A combination of ease of production, algorithms that promote shocking and controversial content, and a lack of moderation have made YouTube in particular a breeding ground for extremist views and racist misinformation.

New Zealand is well placed to fall victim to the hoaxes, pseudoscience and radicalisation that thrive on the Google-owned platform.

New Zealanders now spend more time watching video online on YouTube and Facebook than reading news sites. While Facebook and Twitter have earned the majority of criticism for misinformation, YouTube is likely a more influential source of mis/dis/mal-information.

Compounding this is a suspicion of mainstream news; journalists are the least trusted occupation in New Zealand according to Research NZ, and an underfunded public interest media.

Youtube has pledged to do more to limit the reach of Conspiracy Theory videos: “We’ll begin reducing recommendations of borderline content and content that could misinform users in harmful ways — like videos promoting a phony miracle cure for a serious illness, claiming the earth is flat, or making blatantly false claims about historic events like 9/11” but they have been vague on where they will draw the line and the changes have already been implemented inconsistently. Without a clear policy on what is and is not “borderline content” it is likely there will continue to be harmful content promoted.

The new policy is currently only focussed on a small amount of videos in the United States, but a wider rollout is planned eventually.

## MUCH OF THE NEW ZEALAND-BASED MISINFORMATION ON YOUTUBE SPREADS MISTRUTHS ABOUT NEW ZEALAND HISTORY TO HUNDREDS OF THOUSANDS OF PEOPLE

Searching for something as innocuous as ‘New Zealand History’ on YouTube (in a private browser) brings New Zealand Skeletons in the Cupboard, a discredited racist ‘documentary’ that was pulled from TVNZ for being inaccurate, as the top result. It currently has over half a million views on YouTube.

The YouTube recommendations from that video include a lecture by Jordan Peterson, known for his anti-women views, and other videos about an earlier race of people in New Zealand (6 Reasons New Zealand Was Settled by an Ancient Unknown People, 45,000 views, and Brien Foerster: Crimson Horizon - The Mysterious Red-Haired Sea Kings of the Pacific FULL LECTURE, 38,000 views) as well as videos about Bigfoot.

A video by a New Zealand conspiracy theory channel titled Evidence of Pre-Māori Civilisation has 114,000 views. The same channel has four videos about Southern and Molyneux’s visit to New Zealand.

Within a handful of recommended videos you can find your way from a search for New Zealand history to videos by American far-right figures arguing for white supremacy.

New York Times columnist and Professor at the University of North Carolina Zeynep Tufekci notes that:

*“For all its lofty rhetoric, Google [which owns YouTube] is an advertising broker, selling our attention to companies that will pay for it. The longer people stay on YouTube, the more money Google makes.”*

*“What keeps people glued to YouTube? Its algorithm seems to have concluded that people are drawn to content that is more extreme than what they started with — or to incendiary content in general.”*

The Wall Street Journal conducted an investigation of YouTube content that found that YouTube often “fed far-right or far-left videos to users who watched relatively mainstream news source”. If you searched for information on the flu vaccine, you were recommended anti-vaccination conspiracy videos.

Author J.M. Berger, a scholar with the EU research group VOX-Pol took a quantitative look at users who identify with or are connected to the alt-right movement on Twitter. He conducted a detailed analysis of nearly 30,000 Twitter accounts that either tweeted white supremacist content or followed others who did. The goal was to document the habits and practices of the alt-right on Twitter in a kind of overall “census” of the movement. Berger found that YouTube was, by far, the most common external site white supremacists linked to. There were about 74,000 tweets linking to YouTube videos in his sample; the next-closest external site was Facebook, with only 23,000.

This algorithmic rabbit hole plays a big role in emboldening people to spread race-based online abuse and harassment. Misinformation can create a pretext for racist beliefs and views, and foster a toxic online space where racist language is common and accepted. Online hate speech can lead to the shaping of hateful views and behaviours and, as has been illustrated around the world, violent hate crimes.

## FACEBOOK AND STUFF ARE ALLOWING RACISM TO FLOURISH ON THEIR PLATFORMS

From August to October 2018, ActionStation ran a pilot project called Taiuiwi Tautoko where 20 volunteers were trained and supported to interact with people in online forums for one hour per week to listen, engage and discuss to find common ground with the goal of educating, supporting and encouraging people to develop more caring, thoughtful and educated responses to Māori people, culture and language.

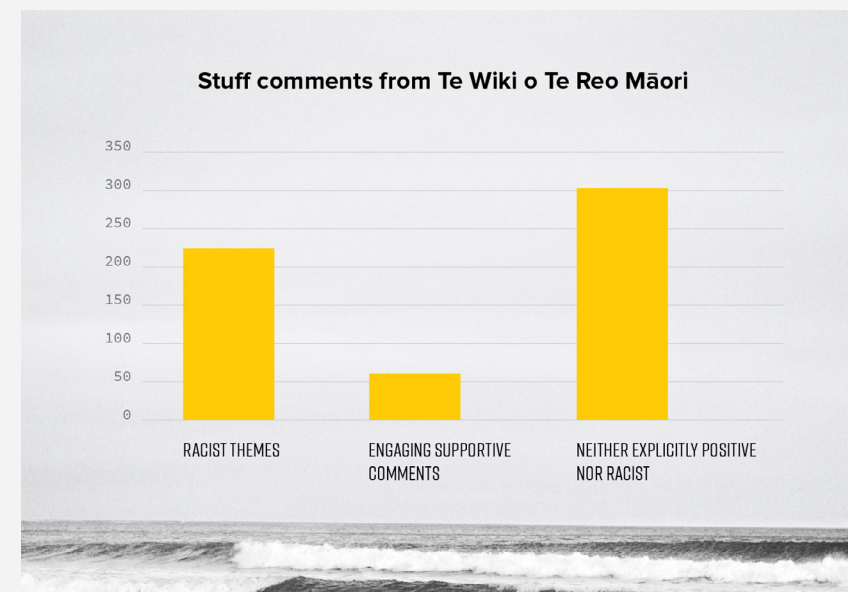
The volunteers detailed their responses and reflected on what worked well (and didn't) in the conversations using Google Forms. Those responses were collated in a spreadsheet that was analysed by an independent researcher from a University of Otago research project (Putting Hope into Action: What inspires and sustains young people's engagement in social movements?).

The focus of the analysis was on the number of racist comments vs caring and supportive comments expressed in two online forums: Stuff and Facebook. Comments were coded against 15 'racist themes' which are detailed in the Methodology section.

Six Stuff articles with a total of 446 comments were analysed for the purpose of this research.

Sixty-six comments were positive and engaging and there were 225 references to the fifteen racist themes. This does not mean that half the comments were racist because each occurrence of a theme was coded and some comments contained material that could be coded under more than one theme. However, the number of references to the racist themes does provide an overall 'feel' for the material encountered in the Stuff comments as the chart below shows. Only 15% of the comments were supportive even when there was a dedicated group of Taiuiwi Tautoko volunteers involved in the online forum.

The six Stuff articles that the Taiuiwi Tautoko volunteers commented on also provides a small database to analyse some trends in the extent and content of racism.



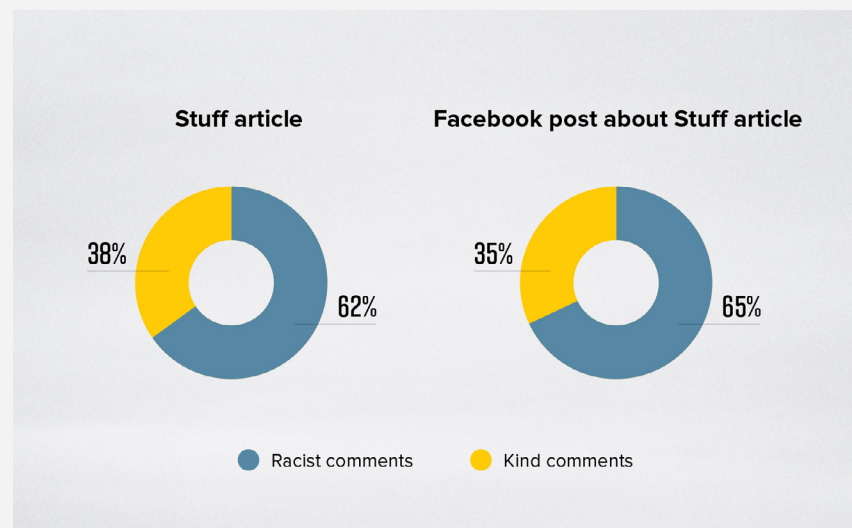
**Our analysis found that, on average, readers of Stuff comments will encounter nearly four times as many different racist ways of thinking for each kind comment. Facebook users will encounter nearly five times as many.**

Eleven Facebook posts were analysed for this research. 4,011 comments were reviewed and coded. As with the Stuff comments, it was possible for Facebook comments to contain text that could be coded across more than one of the racist themes so the number of racist references reported below does not equate to the number of comments. It is a count of the number of times that each racist themes is mentioned. Racist themes emerged 553 times (14%) compared with 124 supportive comments (3%). The majority of comments were neither explicitly positive nor racist at 3,326 (83%). However, these comments should not necessarily be considered 'neutral': they are just neither 'kind nor engaging' nor racist as defined by the fifteen codes. Many of the comments were unpleasant, unkind or at best cheeky.

## WHAT OUR FINDINGS SHOW

The analysis above suggests that comments to both Stuff and Facebook online media are much more likely to contain references to racist themes than comments that are supportive and caring. Additionally, because some posts contain text that can be considered 'racist' in more than one way as identified using the fifteen themes, it was also important to get a comparison using a basic count. This was done using the comments to one Stuff article and one Facebook post (of the same Stuff article).

This basic count shows that in both Facebook and Stuff comments, people are twice as likely to encounter racist comments than kind and caring ones.



ActionStation, along with many other organisations, groups and individuals, are concerned not only that these racist views are held by some people, but that forums exist that allow, invite and support the airing of racist views and attacks. This occurs despite [human rights legislation, and censorship guidelines](#), which explicitly condemn material which “degrades or dehumanises or demeans any person.”

There is [increasing academic](#) and [public unease](#) about internet safety and its implications for society, which have resulted in [guidelines about staying safe online](#) and are annually brought to our attention with “[Safer Internet Day](#)” on the 5th of February. While this safety focus is often directed at keeping individuals safe from breaches of privacy or cyberbullying, [it is increasingly acknowledged](#) that being exposed to racism has serious effects on health and wellbeing. It is therefore crucial that we also address online racial discrimination in order to keep people safe online and promote the type of society that we would prefer to live in.

Both [Stuff](#) and [Facebook](#) have policies about comments but these are not included with the comments sections and have to be searched for. Facebook’s policy was published relatively recently (April, 2018) and hate speech appears to be something Facebook has struggled with, particularly with regard to drawing the line between [hate speech and expression of free speech](#). These boundaries need to be more clearly defined by policy makers and political action so that inciting disharmony and hatred has legal implications.

Finally, what the nineteen different articles/posts and 4,867 comments required for this analysis revealed was that the same names and pseudonyms come up time and time again, which confirms that there are online ‘trolls’. Although there are recommended ways to manage attacks from trolls, it seems incongruent that rather than [restricting trolls](#), the recipients of ‘trolling’ [have the burden of protecting themselves](#).

[Recent research from New Zealand](#) suggests that ‘successful’ trolling requires motivated trolls, reactive targets and the absence of capable guardians. This begs the question: Are Facebook and Stuff capable guardians? Our research would suggest not.

## THE HARMFUL DIGITAL COMMUNICATIONS ACT IS NOT ENOUGH

The Harmful Digital Communications Act (2015) is a powerful piece of legislation that was enacted to address the issue of online abuse. However it is not sufficient to address every issue of online hate, harassment and abuse.

It was passed by the National government to deal with 'cyberbullying', and in some ways it has been successful. Netsafe, the approved agency, reports high success rates in mediation and content removal.

There are, however, some issues.

The law (while broad) is designed for only a limited number of situations where online harm occurs. Specifically, it appears to work well in many cases of one-to-one abuse, where an individual who is being abused can contact Netsafe and identify the abuser.

There have however been instances, some high profile, where seemingly clear cut cases of abuse and harassment are deemed to not breach the act, such as when a Facebook user commented that writer Lizzie Marvelly should try "bungy jumping without the cord".

Netsafe are effective but their work is reliant on the platforms' community standards and enforcement policies to remove content if the abusers themselves won't. Netsafe are not an "enforcement agency." That role is left to the courts.

Unfortunately little support is available for those making HDCA complaints if mediation is unsatisfactory. Netsafe are legally bound to act as impartial mediators, and so are restrained from filling the role of victim support if they recommend a case proceed to court.

There is a concerning rate of drop off between cases that Netsafe recommends proceed to court and those that actually do. This points to a lack of victim support during this process. As well as harming individuals, this contributes to a shortage of case law which would benefit Netsafe.

The tools of the HDCA appear unsatisfactory in other cases of serious abuse online, such as when an organised group (often using 'shell' accounts and fake identities) are targeting an individual. There are also cases where hate is being directed at a group of people, but not necessarily targeted at an individual who can lay a complaint, where there is still a considerable harmful 'bystander' effect.

The HDCA is also an isolated piece of legislation, and a high bar. Most people who face harassment online will never use its mechanisms (though the rate of personal harm complaints is rising year over year).

While the HDCA is an effective tool for many instances of online abuse and harm, there needs to be much more done to address the areas the HDCA fails to cover.

## OUR HATE SPEECH LAWS NEED TO REFLECT AND PROTECT OUR INCREASINGLY DIVERSE COUNTRY

Distinguished Professor Paul Spoonley, Pro Vice-Chancellor in the College of Humanities and Social Sciences at Massey University, says in his [briefing paper](#) about the nature and extent of hate speech in the age of the Internet:

*“The availability and use of online and digital technologies has increased the opportunities for one-to-one or one-to-many communication, including offensive/hate speech...It is therefore timely both to consider the nature of hate speech in New Zealand and to establish the extent and impacts of hate speech on New Zealanders.”*

He notes that there is little systematic recording of the nature and extent of hate speech in New Zealand, partly because there is no agreed definition of what qualifies as hate speech, and there is no requirement for agencies to [systematically collect hate crime or speech data](#). He goes on to say:

*“Germany provides something of a test case for responding to hate speech although it needs to be acknowledged that because of a 20th century history of political and racial extremism, Germany has much stronger limitations on hate speech than most constituencies. Volksverhetzung, or the incitement to hatred, has been a criminal offence in Germany for some time but this was underscored in June 2017 with a new bill concerning hate speech on social media.*

From 1 January 2018, a new German law puts the onus on technology companies to establish the difference between free speech and hate speech. The new law, commonly referred to as NetzDG, requires companies like Facebook and Twitter to remove “obviously illegal” hate speech from its sites within 24 hours of notification or they can be fined up to 50 million euro. As a result Facebook now has 1,200 reviewers based in Essen and Berlin and about 15,000 deletions are made each month in Germany.

In 2017, the UK also announced a new measure to act against hate speech online, and specifically, the Crown Prosecution Service guidelines were revised. These effectively expand the provisions for prosecuting those responsible for hate crimes online.”

In New Zealand, the Human Rights Act currently includes provisions that cover both civil and criminal liability for the incitement of racial disharmony. However, the threshold is extremely high and there is a profound scarcity of successful racial disharmony claims to the Human Rights Review Tribunal.

Racial disharmony provisions only apply to instances where hostility is stirred up amongst people other than those who are the subject of the hate. The expression of hatred in and of itself (or the effect of that hatred on the person or group it is directed towards) is not sufficient for the law to apply. The hate speech provisions in the Human Rights Act also apply only to colour, race, or ethnic or national origins and not religion. ‘Hate speech’ against religion, or even religious people, is not unlawful.

During the most recent [universal periodic review](#), United Nations member states made specific recommendations for New Zealand to improve the human rights of LGBTQI communities including that the government “add gender identity, gender expression or sex characteristics as specifically prohibited grounds of discrimination in Article 21 of the Human Rights Act of 1993” and that “New Zealand amend the Human Rights Act of 1993 to explicitly prohibit discrimination on the basis of gender identity and intersex status”.

Based on the aforementioned Netsafe research that showed people from LGBT communities experience high rates of online hate, these would be useful additions for keeping rainbow people safe on and offline.

## SOLUTIONS COULD LIE IN INDIGENOUS THINKING AND VALUES

“No one’s exercise of free speech should make another feel less free.”

– [Moana Jackson](#)

One of the most significant themes to emerge in this research was the need to attend not just to individualised concerns (e.g. individual rights and privacy) but also to collective dynamics and wellbeing. Therefore any policies that are developed to protect people online and ensure their ability to participate freely and safely online need to have at their centre indigenous and collectivist thinking, especially as Māori have historically (and presently) been among those who are most targeted by hateful speech.

Constitutional lawyer and justice researcher Moana Jackson points out:

*“The right to free speech [in New Zealand]... has too often meant the freedom to hurt, despise, and belittle Māori. Cartoons published a few years ago depicting fat and lazy Māori taking advantage of free school lunches were found by the court to be objectively offensive but protected by free speech and the fact that they weren’t offensive enough to incite a “reasonable person” to hatred or violence against our people.”*

In response to national debate about free speech vs hate speech, Doctoral student in Theology and Religion at Otago University, Graham Cameron writes about the idea of [whakawhanaungatanga](#) instead of censorship.

Whakawhanaungatanga means to build familial connections with others, the relationships that bind us together, our social contract. He argues that defending the rights of racist or sexist people to say whatever they want is implicitly defending their ‘right’ to damage our communities and relationships. Therefore, limiting people’s ability to be sexist and racist is less about censorship and more about upholding whanaungatanga.

To illustrate the difference in the Māori and Western/Pākehā worldviews when it comes to the power of speech, we point to two different whakataukī or proverbs. The first in English: “Sticks and stones may break my bones, but words will never hurt me.” The idea that speech cannot hurt a person. The second in Māori: “He tao rākau, e taea te karo; he tao kōrero, e kore e taea te karo” or “The taiaha can be parried aside but words go straight to the heart.”

In the Māori world, a culture steeped in oral tradition, words create worlds and words matter.

Samoan author Lani Wendt Young, who has received horrific amounts of online abuse, says:

*The public shaming associated with online abuse is amplified many times over when you are Samoan... Because an attack on the individual is an attack on the extended family.*

Our human rights, hate speech and communications laws need to acknowledge collectivist cultures and prioritise collective wellbeing-based solutions.

[Māori digital rights advocate Karaitiana Taiuru](#) says that two Māori values in particular could help support those who build the technology that permeates so much of our lives to build tools for a safer, better internet. Manaakitanga (How can we build tools that encourage users to show each other care and compassion and work to uplift each other?) and Kaitiakitanga (How can we build tools where all users become the guardians of the experience and data in a highly trusted, inclusive, and protected way?).

## LANI'S STORY

Sniper her, rape her, drown her, cook her in an umu' - and Facebook approves

Lani Wendt Young is an award-winning writer, publisher and journalist of Samoan and Māori heritage. She has written a popular series of young adult fiction, is a mother of five, and was named the 2018 ACP Pacific Laureate.

She is also consistently and viciously attacked online.

Death threats, rape threats and threats to her family - the abuse, primarily on Facebook, has made Lani fear for her life.

Lani tried multiple times to get Facebook to remove the abuse but the complaints were not taken seriously because they were in Samoan. Facebook has also said that for a comment to be considered a "credible threat" it has to include specifics: a weapon, time or location.

In April 2018, Lani went to the New Zealand Police with 800 screenshots of violent and threatening abuse. The first officer she spoke to said, "Why don't you just change your name on Facebook so they can't find you anymore?"

They advised her they couldn't do anything about it and she left the police station in tears. In a follow up email Lani was told to stop writing about topics that triggered the harassment.

The abuse began after she testified before the National Commission of Inquiry into Family Violence in Samoa in 2017. Her testimony was about being a survivor of child sexual abuse and recommendations for how to prevent it in Samoa.

While Lani was a journalist reporting on politics in Samoa she wrote articles expressing her sadness and anger about her church's policy on not blessing or baptising children with gay parents or in a same sex marriage.

*There was lots of hate after that.*

That hate turned into an anonymous, targeted and coordinated campaign to frighten, intimidate and silence her.

As Lani said at Netsafe's 2018 Online Safety Conference:

*When seeking to understand Pasifika cultural norms and values in online spaces – it's essential to know that as with other collectivist cultures, we are never just individuals. It's not "me", it's always "us". I speak from the perspective of a Samoan woman, but I believe that the dynamic is very similar for other Pacific Islanders. We are our parents, grandparents, extended family, our village, our church congregations. And when we go online, we carry those connections with us. For a Samoan, using their real name, going online is like being in a village and the internet is a very small place.*

Lani now tries to have someone with her in public at all times. She does not feel safe to travel alone.

Eventually Lani found Netsafe, New Zealand's independent non-profit online safety organisation:

*They filed with the website host provider to have abusive content removed. They were unsuccessful. They filed with Facebook to have abusive content removed. Facebook refused. Netsafe appealed. Facebook agreed and shut down the lead abusers and their anonymous pages. But it was a temporary respite only because the pages appealed and Facebook put them back up within a few weeks, only this time, they were more cocky and assured of their untouchability.*

Facebook appeared to not have a moderator who spoke or understood Samoan to review the abuse Lani reported. It's an example of the irresponsible attitude Facebook (a company worth nearly 500 billion dollars) has had to stopping abuse on the platform. It also demonstrates their failure to scale their moderation and enforcement efforts.

Lani's experience also shows a failure by both the platform and the New Zealand government to keep up with the ways in which online hate is being used as a weapon.

Her experience highlights how even with ample evidence and the HDCA and Netsafe in place, online abuse can be dismissed as unsolvable by law enforcement while new anonymous accounts can be set up on Facebook and the abuse and fear continues.



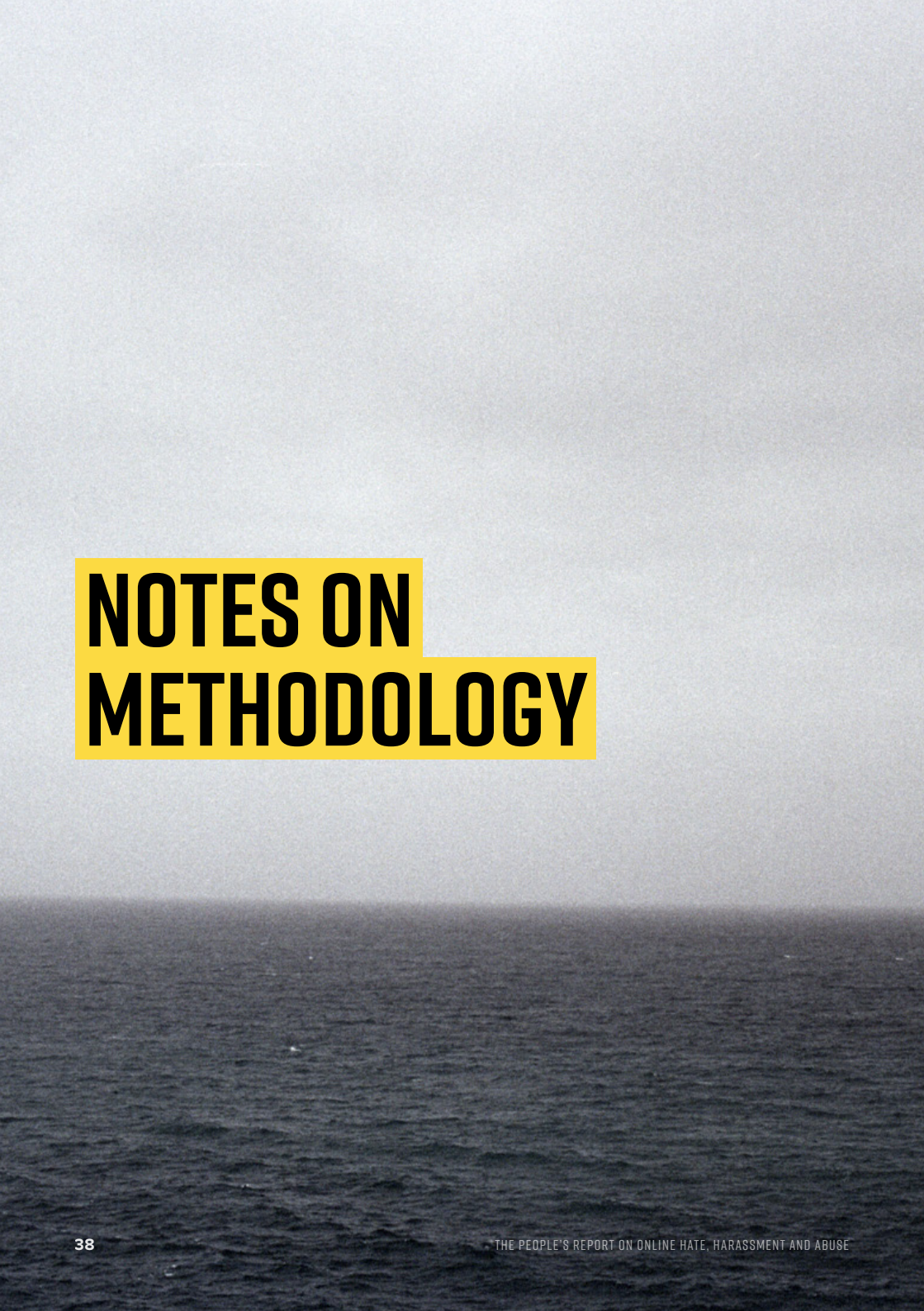
# RECOMMENDATIONS

## RECOMMENDATIONS

If the internet is the new public square, it is imperative that lawmakers ensure the ability of all New Zealanders to access reliable and credible information about issues of public importance, and the ability of everyone in this country to participate safely in public conversations about those issues.

Based on our analysis, we are making four recommendations to the New Zealand government:

- 1. Remove:** Ensure platforms are active in removing harmful content quickly. An investigation into the most effective method to do this would be required, but the responsibility should be placed on the platform, not the users.
- 2. Reduce:** Limit the reach of harmful content. Neither the platforms nor the users who create hateful and harmful content should benefit from algorithms that promote divisive and polarising messages.
- 3. Review:** The New Zealand government needs to review our hate speech laws, the Harmful Digital Communications Act, the Domestic Violence Act, the Harassment Act and the Human Rights Act to ensure they are fit for purpose in protecting people online in the 21st century.
- 4. Recalibrate:** One of the most significant themes to emerge in this research was the need to attend not just to individualised concerns (e.g. individual rights and privacy) but also to collective dynamics and wellbeing. Any policies that are developed to protect people online need to have indigenous and collectivist thinking at their centre. They should also ensure that all internet safety / hate speech agencies funded by the Crown reflect the increasing diversity of our country.



# NOTES ON METHODOLOGY

## NOTES ON METHODOLOGY

### UMR SURVEY

---

We commissioned UMR Research to conduct an online survey of people of colour in New Zealand to measure the extent and impact of racialised online abuse.

The survey asked 618 Māori, Pacific and Asian people their experiences, and the results were weighted by age, gender, and ethnicity.

The breakdown of respondents is as follows: Māori (n=210), Pacific (n=205) and Asian (n=203) respondents. The survey asked about social media usage, experiences of online harassment and the impact of that harassment.

Fieldwork was conducted from the 8th to the 20th of August 2018.

## STORY COLLECTION

We also used an online form to gather personal stories of online harassment. We asked 12 questions covering the context of the abuse, platform, and the abuser. We had 69 in depth responses from around New Zealand.

The limitations of this are substantial. Those driven of social media by harassment are significantly less likely to have seen and taken part in the survey, meaning people facing the worst cases of harassment likely do not have their voices represented here.

The experience of serious harassment is traumatic, and this likely discouraged some who have experiences to relive them by taking part in the survey. On average it took people 15 minutes to complete the survey, meaning it took a significant amount of time, which likely discouraged some people with detailed experiences.

The full text of the survey form we used to solicit stories from individuals with experience of harassment online is below.

1. Did you experience the abuse, witness it or both?  
*Experienced abuse • Witnessed abuse • Both*
2. On which website or platform did you experience or witness abuse and harassment?  
*Twitter • Facebook • Instagram • YouTube • WhatsApp*
3. What was the context of the abuse? E.g. did it happen in the context of an online discussion of a specific topic?
4. Did the abuse and harassment happen in a private message, in a public comment (e.g. in a comment thread on Facebook or Twitter), in both or something else?  
*Public • Private • Both*

5. What do you know about the person/account responsible for the abuse? E.g. was it one person, or more than one? Were they someone who you know, someone who you don't know but who was easy to identify, or someone anonymous. Did you have reason to suspect the abuse was being generated by an organised attack? Or reason to think the attack was being made by a bot or bots? If so please tell us more about that.
6. What forms did the abuse and harassment take? E.g. direct or indirect threats of violence, sexist, sexually explicit, homophobic, transphobic, ableist, or racist language, comments or images (and whether or not they were directed at you or not), comments or images that insult a particular religion or culture, other abusive language, comments or images, or posting your personal details with the aim to cause you alarm or stress?
7. Do you think any of the following had to do with the online abuse you faced?  
*Your ethnicity • Your gender • Your sexual orientation • Your religion  
Your political affiliation • Your job • None of the above*

[Logic jump if A - F chosen above]

8. Could you tell us more about that?

Thank you, now we want to ask about the impact of this online abuse and harassment.

9. How did the abuse and harassment affect you emotionally, psychologically and socially? E.g. did it make you feel that your own safety or the safety of a family member or friend was threatened, that your job or job prospects may be threatened, were you less able to focus on everyday tasks, to concentrate, to make day-to-day decisions, did it affect your ability to sleep, or to accomplish what you usually would in a day. Did you experience panic attacks, anxiety or stress, a feeling of apprehension when receiving emails or social media notifications, lower self-esteem or a loss of self confidence or mood swings. Did you feel powerless, isolated?

10. Since experiencing abuse/harassment on social media, has your usage of those social media platforms changed? E.g. have you stopped using the social media platforms, used them less, blocked users who were responsible for the abuse or harassment, made your account private or protected, increased the privacy or security settings on your account, or changed your use of social media in any other way?
11. Since experiencing abuse/harassment on social media, has the type of content that you posted on those social media platforms changed? E.g. have you stopped posting or sharing content that expresses your opinion on certain issues, and if so, which issues? Have you stopped posting or sharing some other kind of content, and if so which kind? How else have you changed the content you post on these social media platforms.
12. What needs to happen? Based on what you have seen, heard or read, what response would you like to see from the Government, Parliament, the police and the tech companies (e.g. Facebook, Twitter and Google) to addressing online abuse and harassment in New Zealand?

## PREVALENCE OF RACISM ON STUFF AND FACEBOOK COMMENTS ANALYSIS

Racism has been defined in different ways at different times by different disciplines but at the core of racism is 'an ideology of racial discrimination' (The Bridge over the Racial Divide: Rising Inequality and Coalition Politics by Wilson, W.J.). In her [report to the Race Relations Commissioner](#), Jenny Rankine defines racism as:

*'An ideology of superiority embedded in powerful institutions and social norms that create and maintain avoidable and oppressive systems of inequality between dominant and other ethnic and cultural groups.'*

Rankine's report outlines the history of racism in Aotearoa New Zealand noting the impact of colonisation, historical discrimination against Māori and other cultural groups (particularly Chinese, Indian and Pasifika immigrants) and how particular historical events can suddenly alter the climate around racism. Examples in New Zealand include the 1974 'dawn raids' looking for illegal Pasifika immigrants, the 1981 Springbok tour and Don Brash's 2004 Orewa speech. As time goes by, and as these and other events influence what is acceptable to say and do, the nature of racism has changed to be more complex. Some argue that as racism is increasingly acknowledged and called out, Pākehā are increasingly anxious about their place in Aotearoa and [resort to more subtle forms of racism](#) to maintain the status quo.

This 'modern racism' is just as harmful as the most obvious forms. Rankine states that ethnicity is less likely to be mentioned in new racist arguments. Instead contemporary forms of racism defend dominant values, exaggerate cultural differences, deny that serious racism or prejudice still exist and claim that inequality is a result of individuals' own actions. This type of racism ignores the processes, structures and systems that are instigated and valued by the dominant group. Overt racism is increasingly taboo and unlawful under the [Human Rights Act](#). As a result denial of racist comments, actions or systems has become another way of protecting and defending white privilege. The analysis of the Tauiri Tautoko project data presented in this report was informed by these understandings of racism.

Because of the extent of data involved in the Tauwi Tautoko data project, which is not straightforward (responses to articles may include comments on the articles, responses to comments, exchanges between individuals, likes, dislikes, moderator removal of some comments) analysing the data was a challenge. The decision was made to focus on only one week and provide a picture of the extent and nature of the type of online comments. The week we chose to focus on is September 10th until 16th. This was te wiki o te reo Māori so there was particular interest in and media reports about Māori language and culture during this week.

The coding of the data was informed by Rankine's report described above as well as [Kupu Taea: Media](#) and te Tiriti Project and Max Harris's "The New Zealand Project." From these sources fifteen themes were developed and comments were coded as racist if they include any of the following characteristics:

- Compares Māori culture to dominant Pākehā or other cultures, stating or implying that it is inferior, less important, wrong, irrelevant or insignificant.
- Uses stereotypes and/or exaggerations to describe Māori culture, individuals or groups which reduces and simplifies complex and variable ideas, customs and behaviour.
- Conflates Pakeha values, norms and mores with New Zealand values, norms and mores and/or states that Māori values, norms and mores are irrelevant or not useful to NZ and New Zealanders/self.
- Deny that racism exists or that it is insignificant compared to the past or in other places or groups in the population (such as the elderly or conservative groups).
- Suggest that inequalities or differences are the fault of minority cultural groups and it is their responsibility to keep their culture alive. Also includes comments about the majority of Māori not being interested in their culture or language.
- Deny that historical injustice to Māori occurred or that it is not as bad as it could have been or not as bad as it is in other countries or that is how things were done back then
- Claim that inequality does not exist or that everyone has equal opportunities.

- Criticise or judge minority cultures for not adhering to the dominant Pākehā values, behaviours or views.
- Describe Māori cultural values or perspectives as threatening or challenging to the accepted way of being or doing things or are imposing their values on others. Includes stating that addressing Māori needs and rights is undemocratic, not in the interest of the majority or impinges on others' human rights.
- Deny white privilege and state that it is actually Pākehā who are the victims of racism, discrimination or political correctness.

Purport that everyone has equal opportunities so any demands or requirements for special treatment or reparation for previous harm are unjustified.

Contend that Māori have already received enough additional support and reparation for historical injustice is complete or that it's impractical or impossible to right past wrongs and that any claims or demands by Māori are unjustified, greedy or money-seeking.

Question whether Māori are the tangata whenua of Aotearoa/New Zealand with strong spiritual, cultural and historical connections to this land.

Maintain that we are all kiwis, New Zealanders, immigrants or 'mixed race' and that acknowledging cultural difference is divisive or intended to cause trouble.

Stating that te Tiriti o Waitangi is irrelevant and/or blocking future development.

The criteria required for a comment to be included in the positive count comments needed to engage in a thoughtful, caring, open, and inquisitive way and provide some re-education if it was relevant or welcome because that was the goal of the Tauwi Tautoko project.

Facebook comments were more difficult to code than the Stuff comments because often Facebook commenters know each other and there are personal exchanges and accusations or humorous exchanges. Sometimes it is difficult to read the tone of the comments, particularly if they are sarcastic, personal or responding to deleted comments.



# ACKNOWLEDGEMENTS

## ACKNOWLEDGEMENTS

This report was written by Laura O'Connell Rapira and Leroy Beckett from ActionStation.

ActionStation is an independent, crowdfunded, community campaigning organisation. Our mission is to tautoko (support) and whakamana (uplift) everyday New Zealanders to act together in powerful and coordinated ways to create what we cannot achieve on our own: a society, economy and democracy that serves all of us - everyday people and Papatūānuku.

Marianne Elliott, Co Director of The Workshop, helped design the research questions.

Judith Sligo is a Senior Research Fellow in the Department of Preventive and Social Medicine at the University of Otago. She is a Pākehā university researcher from the South Island with 20 years qualitative and mixed methods research experience, primarily in the areas of health and wellbeing of children, young people and families. She conducted the analysis of our Tauīwi Tautoko project data.

Renee Pearson assisted with visual design.

PEOPLESCHARASSMENTREPORT.COM

---