

A REPORT BY THE WORKSHOP

2019

MAY

D I G I T A L
T H R E A T S
T O
D E M O C R A C Y

The
Workshop

The
Law
Foundation
NEW ZEALAND

Luminate

D I G I T A L T H R E A T S T O D E M O C R A C Y

This report is part of the Digital Threats to
Democracy research project.

To see the rest of the reports and the
overall findings go to digitaldemocracy.nz

ISBN: 978-0-473-48026-4

CONTENTS

ACKNOWLEDGEMENTS	05
INTRODUCTION	06
EXECUTIVE SUMMARY	10
PURPOSE, DEFINITIONS AND METHOD	29
Research purpose	29
Definitions	29
Methodology	30
Overview	32
THE OPPORTUNITIES OF DIGITAL MEDIA	35
Democratization of information publishing	36
Broadening of the public sphere	36
Increasing equality of access to and participation within political processes	36
Increasing engagement in political processes	36
Increasing transparency and accountability from government	36
Promotion of democratic values	37
THE THREATS	38
Increasing power of private platforms	39
Foreign government interference in democratic processes	39
Surveillance and data protection	40
Fake news, misinformation and disinformation	40
Filter bubbles and echo chambers	41
Hate speech and trolling	41
Distrust/dissatisfaction with democracy	42
THE SOLUTIONS	43
A Hierarchy for Solutions and Interventions	43
Change society wide structural & systems issues	46
Create supportive environments & contexts - making the default inclusive and safe	47
Create long lasting protections for people or intervene to protect them	49
Build understanding and change individual behaviours	50
WHAT ROLE FOR NZ?	51
New Zealand as follower	52
New Zealand as leader	52
New Zealand as niche influencer	52
CONCLUSIONS	53
Inclusive and transparent processes are critical	54
More research needed	54
Evidence-led and principled approach to urgent policy development	54
Apply human rights principles	55
Agile and responsive approach to policy	55
Urgent areas for action	56
SUMMARY OF KEY FINDINGS FROM EACH PART OF THE RESEARCH	57
Key findings from the survey	57
Key findings from the literature review	58
Key findings from the interviews	59

CONTENTS

APPENDIX 1: LITERATURE REVIEW PART 1	65
APPENDIX 2: LITERATURE REVIEW PART 2	93
APPENDIX 3: REPORT ON QUANTITATIVE SURVEY	123
APPENDIX 4: REPORT ON QUALITATIVE INTERVIEWS	143
APPENDIX 5: ONLINE HATE AND OFFLINE HARM REPORT	227

ACKNOWLEDGEMENTS

This research was made possible by support from our major funder, the New Zealand Law Foundation's Information Law & Policy Project, with additional research funding from the Luminate Group.

The New Zealand Law Foundation - Te Manatū a Ture o Aotearoa – provides grants for legal research, public education on legal matters and legal training. The **Information Law & Policy Project (ILAPP) funds** projects that will better prepare New Zealand for the challenges of the information age. The project is intended to support the growth, understanding and resilience of New Zealand and prepare the country for future digital competence. It will support and feed into work the public and private sector is undertaking, but will remain independent.

The Luminate Group is a global philanthropic organisation with the goal of empowering people and institutions to work together to build just and fair societies. They fund and support projects that will help people participate in and shape the issues affecting their lives, and make those in power more transparent, responsive, and accountable. Their focus is on civic empowerment, data & digital Rights, financial transparency, and independent media.

The research team on this project was lead by Marianne Elliott (The Workshop) and included Dr Jess Berentson-Shaw (The Workshop), Dr Kathleen Kuehn (Victoria University of Wellington), Dr Leon Salter (Massey University) and Ella Brownlie (The Workshop). Project management was provided by Jay Brooker (The Workshop). The quantitative survey was conducted by UMR Market Research.

We particularly want to thank all the participants in the interviews for this research. They volunteered their time on this project, brought both goodwill and clear thinking to the unreasonably broad scope of the research, and responded rapidly to our requests for feedback. This kind of generosity and rigour across the sector gives us hope that with more investment in research and a more joined-up government approach to the subject, New Zealand will be able to contribute substantially to an urgently needed global collaboration on solutions.

INTRODUCTION

In February this year, as I pored over the findings of our literature review and read through hundreds of pages of interview transcripts, I wrote that an adequate response to the problem of online hate, harassment and abuse was possible. It would require a recalibration of our policy approach, some international diplomacy and cooperation, and a sufficiently diverse group of decision-makers at the helm.

I believed then that all of that was within the capacity of the New Zealand government, and that there was “likely to be a leadership role for our country in global efforts to combat online abuse and, as Sir Tim Berners-Lee has put it, ‘fight for the web’.”

This belief has proven to be founded, although under circumstances none of us ever wanted to witness. As we completed this research, it was announced that New Zealand’s Prime Minister, Jacinda Ardern, would meet the French President, Emmanuel Macron in Paris to “bring together countries and tech companies in an attempt to stop social media being used to promote terrorism.” The meeting will invite world leaders and tech company CEOs to sign a pledge called the ‘Christchurch Call’.

In many ways, the devastation of the Christchurch mosque massacres has proven to be a turning point for New Zealand on this, and other policy issues. We now know that our small size and relative remoteness do not render us immune to the terrible harm that can be done by a person motivated by hatred, inspired by the internet and armed with a semi-automatic weapon.

In the wake of the March 15 attacks, in response not only to the unthinkable cruel and manipulative use the terrorist made of the internet in the course of the attack but also to the many ways in which online spaces have allowed hatred to grow and spread, many people - including our Prime Minister - called for greater accountability and care from the big digital platform companies, including Facebook and Google.

It’s a call some of us have been urging our government to make for some time now, and many feel it is long overdue. But here we are now, and this is a crucial moment in the history of the relationship between citizens, governments around the world and a handful of people who not only control a significant portion of the means by which we all communicate and the distribution of news and information to vast percentages of the world’s population, but also hold huge quantities of personal data about us all.

The question is no longer whether something needs to change. The question has become: what precisely needs to change? And even more importantly: what can be done? What evidence do we have as to the interventions and solutions that might mitigate against the biggest threats posed to our democracy by digital media, without losing the best of the opportunities that the internet offers? Those are the questions we set about answering with this research.

We are far from the first people to tackle these questions, as our literature review reveals. Researchers, academics, journalists and former employees of the big tech companies have been studying and writing about the impact of digital media on democracy in increasing numbers over recent years.

In his book, *The People vs Tech* Jamie Bartlett predicted: “In the coming few years, either tech will destroy democracy and the social order as we know it, or politics will stamp its authority over the digital world.” In his view, “technology is currently winning this battle, crushing a diminished and enfeebled opponent.”

Similarly, in *How Democracy Ends*, David Runciman assessed the comparative strengths of the tech giants versus governments, in a ‘Leviathan vs Leviathan’ showdown for the future of democracy. Although he gave governments more of a shot than Bartlett had, he concluded that while “Facebook will not take down the Leviathan in mortal combat ... it could weaken the forces that keep modern democracy intact.”

But neither Runciman nor Bartlett, nor any of the analysis I’ve read over the past year, predicted the situation we are now in. None imagined a Prime Minister with a global reputation for compassion, armed with moral courage, clarity and the support of an outraged nation.

Has Jacinda Ardern become the global leader capable of taming the tech giants? There are good reasons to hope so, and even more reasons to ensure that this rare opportunity is neither wasted nor lost.

NEED FOR A COHESIVE, EVIDENCE-BASED APPROACH TO POLICY

One of the challenges of rapidly developing policies on digital media in response to a situation like the Christchurch attacks is that this entire area of policy has been relatively neglected until recently. As one interviewee in this research said, we need a better system for making policy on these issues before we can be any kind of global leader. “Smart people just basically giving their opinion with no real information behind it,” won’t be good enough to develop the kind of solutions demanded by this particular set of problems and, they say “that’s how we’ve made our policy in this space, generally.”

Until very recently, there was no centralised or coordinated government process for developing policies and strategies in response to the challenges posed by digital media. Responsibility fell to a wide range of different agencies and teams, and policy development was consequently, inevitably, fragmented. In the process of doing this project, we found it difficult to establish who in government, if anyone, had a broad view over the full range of issues raised in our research. Recently, new efforts at coordination have begun to appear, with some degree of overarching responsibility, although not necessarily with the resources needed to develop policy across such a wide-ranging and rapidly changing area.

In the past, according to one interviewee, New Zealand has either simply adopted the policy approach taken in another jurisdiction “or we have a relatively flimsy policy discussion which isn’t founded in evidence.” In order to build our capacity as a country to understand and deal with these issues, they argue, we need more of an evidence base. “Before we can be leaders in any sense, we need to be equipped to have a solid base for developing policy ourselves.”

What our research shows is that it is critical that the Prime Minister and her advisors look beyond immediate concerns about extremism and content moderation, and ensure that our government's efforts in this moment take into account the wider structural issues that created the conditions in which a live video of an act of such violence could be shared and viewed so widely.

Those wider structural issues include in particular the impact of **platform monopolies**, in which a handful of people have the power to determine the social interactions and access to information of millions of people, **algorithmic opacity**, in which algorithms have ever-increasing influence over what we hear and see without appropriate transparency or accountability, and the **attention economy**, which gives priority to content that grabs attention, without sufficient regard to potential harm.

Our intention is that this research will contribute to a wider consideration of the issues arising from digital media's impact on democracy, and to the development of a body of evidence which supports this critical work.

Marianne Elliott

Lead Researcher, Digital Media and Democracy

Co-Director, The Workshop

A REPORT BY THE WORKSHOP

2019

MAY

D I G I T A L
T H R E A T S
T O
D E M O C R A C Y
E X E C U T I V E
S U M M A R Y

The
Workshop

The
Law
Foundation
NEW ZEALAND

Luminate

D I G I T A L T H R E A T S T O D E M O C R A C Y

This report is part of the Digital Threats to
Democracy research project.

To see the rest of the reports and the
overall findings go to digitaldemocracy.nz

ISBN: 978-0-473-48026-4

INTRODUCTION

As we completed this report it was announced New Zealand's Prime Minister, Jacinda Ardern, would meet the French President, Emmanuel Macron in Paris to "bring together countries and tech companies in an attempt to stop social media being used to promote terrorism." The meeting will invite world leaders and tech company CEOs to sign a pledge called the 'Christchurch Call'.

The question is no longer whether something needs to change. The question has become: what precisely needs to change? And even more importantly: what can be done? What evidence do we have as to the interventions and solutions that might mitigate against the biggest threats posed to our democracy by digital media, without losing the best of the opportunities that the internet offers. Those are the questions we set about answering with this research.

One of the challenges of rapidly developing a policy response on digital media in response to a situation like the Christchurch attacks is that this entire area of policy has been relatively neglected until recently. As one participant in this research said, we need a better system for making policy on these issues before we can be any kind of global leader. In order to build our capacity as a country to understand and deal with these issues, we need a better evidence base.

What our research shows is that it is critical that the Prime Minister and her advisors look beyond immediate concerns about violent extremism and content moderation, to consider the wider context in which digital media is having a growing, and increasingly negative, impact on our democracy.

BACKGROUND

Over recent years a growing body of international research has looked at the impact of digital media on democracy, with particular focus on the US and the UK, where the role played by digital media in the election of Trump and the Brexit referendum raised significant concerns.

Our project was designed to find out if we should be worried about these same issues here in New Zealand, and if so, what should we do about it? In order to answer that question we identified five key features of democracy against which we could measure the impact of digital media, for better and for worse. They are:

- Electoral process and pluralism
- Active, informed citizens
- Shared democratic culture
- Civil liberties and competitive economy
- Trust in authority

WHAT WE'VE FOUND

Critically, we found that digital media is having an impact across every one of those features of a healthy democracy.

There are indicators that digital media has had some beneficial impacts. Our quantitative research here in New Zealand indicates, for example, that people from minority groups have been able to use digital media to participate in democratic processes including accessing political players, and engaging in public debate. Whatever our response to the challenges posed to democracy by digital media, it's important we don't lose these opportunities in the process.

But the overall trend should raise serious concerns. Active citizenship is being undermined in a variety of ways. Online abuse, harassment and hate - particularly of women, people of colour, queer people, people with disabilities and people from minority religions - undermines democratic participation not only online, but offline. Misinformation, disinformation and mal-information are undermining not only informed debate, but also public trust in all forms of information. Distraction and information overload are eroding citizens' capacity to focus on important and complex issues, and their capacity to make the 'important moral judgements' required in a healthy democracy.

Likewise, interviewees described a myriad of ways in which our shared democratic culture is being undermined by digital media - including through disinformation, polarisation, attention hijacking and radicalisation.

One of the clearest impacts of digital media on our democracy has been its impact on funding for mainstream media. While Facebook and Google hoover up the advertising revenue that once would have been spent on print, radio and television advertising, they contribute nothing to the work of producing the kind of news and current affairs reporting that is essential to a functioning democracy.

The representative survey we carried out indicates that New Zealand's small size and relatively healthy mainstream media (relative to elsewhere and despite significant resource

challenges) may help us avoid the worst effects of ‘filter bubbles’ and ‘echo chambers’ in digital media on some issues.

Interviewees in our qualitative research nonetheless pointed to examples where debate in New Zealand about issues like free speech, hate speech and gender identity attracted the attention of foreign actors holding strong, even extreme, views on these issues.

Engagement by these foreign actors in the online public debates on these issues here in New Zealand appears to have contributed to a polarisation of views here.

THE THREE CORE PROBLEMS TO EMERGE FROM OUR RESEARCH

At the heart of the challenges to democracy posed by digital media are three core problems:

1. Platform monopolies: two or three corporations control not only our means of communication, but also the content which is distributed, both of which are core aspects of our democracy. Whilst the market power and global mobility of these companies make it possible for them to avoid national regulatory measures, either by moving operations elsewhere or simply ignoring them;
2. Algorithmic opacity: algorithmic engines are using huge quantities of personal data to make ever more precise predictions about what we want to see and hear, and having ever increasing influence over what we think and do, with little transparency about how they work or accountability for their impact; and
3. Attention economy: the dominant business model of digital media prioritises the amplification of whatever content is best at grabbing our attention, while avoiding responsibility for the impact that content has on our collective wellbeing and our democracy. The negative impact is brutally clear from both the literature and the world around us.

THE NEED FOR A SYSTEMIC RESPONSE

The key message is clear; digital media is having massive, system-wide impacts on our democracy. It affects every part of our lives and the people who run the corporations controlling the major platforms are having a determinative impact on the very structures and functions of our society. While better content moderation is clearly one of the responses we must demand of the platforms, it is not even close to being a sufficient response to the scale of the challenge.

It’s critical that this moment of global cooperation is used to address the wider, structural drivers of the biggest threats posed to democracy by digital media. These structural drivers include the power that a handful of privately-owned platforms wield over so many aspects of our lives, from what information we see, who we interact with, and who can access information about us. And we must do this while maintaining and building upon the many opportunities digital media simultaneously offer, to tackle some of the biggest challenges facing democracy, including inequity of access and declining engagement.

THE OPPORTUNITIES OF DIGITAL MEDIA

The potential of digital democracy lies in its ability to increase democratic participation, embrace diversity of opinion, and empower marginalised groups. We identified six clear opportunities from the literature that digital media offers. These are: the democratisation of information publishing, broadening the public sphere, increasing equality of access to and participation within political processes, increasing participation and engagement in political processes, increasing transparency and accountability from government and promotion of democratic values.

Broadly speaking these opportunities fit into two categories: First, those that enable individuals, citizens, or groups, who due to their status in society have been excluded from fully participating in different aspects of the democratic process, through greater access to the levers of democracy. Examples include the use of digital media to: broaden the public's engagement with indigenous people and their lives, give more exposure to women in politics, build well-networked, educated and empowered communities, and encourage political engagement from youth.

The second category of opportunities relate to digital media's use by people in governments to make processes of democracy more inclusive, to increase engagement with citizens, improve transparency of government work, and rebuild trust in democratic processes. Examples of such work include: online deliberative democracy processes, open or e-government initiatives, and funding of public service journalism, platformed on digital media.

The opportunities for digital media are significant and important. If used well, digital media can enable governments to respond effectively to the experiences of marginalised groups, to ensure equitable policies and practices are designed, delivered and adjusted, and to build trust in the democratic institution as responding to the needs of all people. It offers as much to people pushing against barriers to their progression, inclusion, and improved wellbeing in society, as it does to people in government looking to remove those barriers and build a more inclusive democratic system.

THE THREATS

The threats to this promise outlined in the literature are significant however, and most are intricately bound up with the concentration of power in profit driven companies. The seven key threats we identified to inclusive democracy from digital media were: the increasing power of private platforms, foreign government interference in democratic processes, surveillance and data protection issues, fake news, misinformation and disinformation, filter bubbles and echo chambers, hate speech and trolling, and distrust/dissatisfaction with democracy.

Researchers highlighted the increasing dominance of an increasingly small number of privately-owned platforms over the internet. People who own and control these platforms have a monopolisation tendency linked to the relationship between the mining of user-data and their imperative to make profits. This model of operation is termed “platform monopolies”. The monopolisation tendency makes alternatives to the data-extraction for profit model, for example, co-operative, democratised ownership models, hard to start up and survive.

The concentrated power of these platforms shapes not just the wider information context and ability to develop alternative non-extractive models of digital information provision and sharing, but individual’s personal lives also. Platform monopolies affects how we interact socially and with whom through algorithms. A body of literature points to how the actions of the people running these companies impact human rights, both through the control of personal data and the level of control over what appears in the public sphere.

From this model of platform monopolies flows a series of further threats to democracy. Some relate to the features of the platforms, directly linked to the capture of people’s personal data. The collection and on-sale of personal data by these platforms, to both governments looking to undertake surveillance on their own citizens, and private organisations wishing to make profits, erodes trust in information systems by the public, and curtails the professional work of media and writers - a key plank in our democracy.

The creation of the “attention economy” also poses a significant threat. People’s propensity to attend to shocking, false, or emotive information, especially political information, is exploited and used as a commodity product by digital media platforms. The literature shows that governments with the means and inclination to manipulate information can tailor false information towards individuals with the express intent of interfering in other countries democratic processes, for example the Russian government interference in the US election of 2016 using ‘bots’ and disinformation campaigns. While misinformation & disinformation, especially political disinformation, targeted at

individuals on digital media, is used to influence politics with a big and small p, from national elections through to information provision and sharing with regard to political issues and policy more generally.

Filter bubbles are a specific technical effect of this attention economy. Facebook's news feed is a filter bubble, created by a machine-learning algorithm which draws on data created by user networks, likes and comments, and how much organisations are willing to pay to be present there. Filter bubbles and the related echo chambers they feed into (in which people attend only to information which confirms what they already believe) are linked to a decline in trust in the ability of traditional news media to provide reliable information, have been found to exacerbate political divisions and polarisation, and has negative implications for the mechanisms of liberal democracy, as developing a broad consensus around decisions made in the public good becomes increasingly difficult.

The rise of hate speech and trolling is linked to the polarising effects of filter bubbles and echo chambers. Both racialised and sexualised hate speech is a specific threat exacerbated by the anonymity provided by digital media. Hate speech and trolling on digital media encourages affected groups to retreat to safe locations, rather than engaging with national debates and institutions. Research has found a correlation between strong, vocal disagreements with an individual's perspectives and a "spiral of silence" which acts to curtail the voicing of contentious opinions by minority groups. The particular ability of trolls and hate speech to fan antagonistic 'flames' rather than promote rational debates, has a direct impact on democratic participation.

While people's distrust with democratic process is a longer term issue, digital media has likely exacerbated this pattern across western democracies. Researchers argue that trust, informed dialogue, mutual consent, and participation - fundamental features of democracy - are being eroded by the features that make social media so profitable. Researchers also found that the way in which the information is distributed on digital media (horizontal, and decentralised and interactive) increases intolerance of others, polarisation and skepticism toward democracy.

The opportunities of digital media, while still apparent, appear to have been suppressed by the sheer weight of fake news, filter bubbles, populism, polarisation, hate speech, trolls, and bots, that have emerged from the concentration of power in a small group of private organisations seeking to maximise profits. Digital platforms initially celebrated for their democratic possibilities, have transformed into anti-democratic power centres through the collection and exploitation of users attention and data. These privately owned platforms have largely escaped public oversight or regulation over their ability to harness this new power for commercial or political gain.

The question is what can policymakers do to recalibrate? Are there empirically tested public policies and approaches that can ensure digital media works to strengthen and deepen democracy?

THE SOLUTIONS

While our literature review was not exhaustive, in general we found a dearth of empirically tested solutions. Likewise, and possibly because of the lack of evidence in the literature, the experts interviewed for this research generally had more to say about the risks and threats they saw arising from digital media, than they did on potential workable solutions. However, we did find some common ground on solutions between the literature and in the interviews. Below we discuss potential solutions for each of the identified threats, with a focus on optimising the opportunities. We have organised those solutions in a hierarchy, from those we think will have the greatest impact, with the least effort required by individual citizens, through to those with the least impact and most individual citizen effort. That is not to say that the politics of implementing those solutions with the greatest impact are not difficult, but that the political effort required is justified by the potential for positive impact.

CHANGE SOCIETY WIDE STRUCTURAL & SYSTEMS ISSUES

This section focuses on structural and systemic change, addressing for example the disproportionate power of the tech giants vis-a-vis governments, citizens and their domestic competitors.

REDUCE THE POWER OF PRIVATE PLATFORMS BY:

Regulating platforms like other industries. Currently, regulatory debates largely centre around defining the structure, terms and conditions of what kind of industry private intermediaries represent. How platforms should be regulated or governed thus partly hinges on how these services are defined; for example, whether social media platforms are media companies, public spaces, utilities or some other service, largely informs how they can ultimately be governed. There is little or no empirical evidence to show how regulation in this area would or would not work, and therefore adaptive approaches to policy and regulation will be needed. This will involve ensuring that the impacts of

any change are regularly monitored and changes made as needed in response to those findings.

Introducing new modes of collective action. Under industrial capitalism we had collective bargaining, the strike, for example forms of collective action that were sanctioned by law and had the support of society that allowed people to tame capitalism with legal protection. In relation to digital media researchers suggest there are opportunities for more collective action both by tech workers, demanding for example more ethical design in the products they work on, and by digital media users.

COMBAT FAKE NEWS BY:

Supporting a vibrant and diverse media sphere. One that balances strong, independent and adequately resourced public service media with a non-concentrated commercial media sector. This is proposed but untested idea.

CREATE SUPPORTIVE ENVIRONMENTS & CONTEXTS - MAKING THE DEFAULT INCLUSIVE AND SAFE

REDUCE THE POWER OF PRIVATE PLATFORMS BY:

Designing new competitive digital media solutions. Disruptive technology is needed to forge an alternative digital future that in turn, facilitates a more democratic internet. This means the creation of platforms offering a different set of affordances (ie not those driven by platform monopolies).

REDUCE INTERFERENCE FROM FOREIGN GOVERNMENTS AND POWERS BY:

Designing new cybersecurity infrastructure and drawing on “big datasets” to review and assess electoral policies. The research in this area is also largely normative, and seems to generally prescribe such infrastructure and reviews will reduce threats to elections and other political processes.

ADDRESS SURVEILLANCE & IMPROVE DATA PROTECTION BY:

Regulating companies’ information management practices. Some regulatory measures, like the Singaporean Data Protection Act 2012, work to and have been proven effective in bringing formal charges to data mismanagement and abuse.

Making regulatory changes to data privacy policies. However, there is little evidence to suggest that these changes will reduce surveillance/data collection so much as regulate how that data is stored, accessed and used by data collectors and other third parties

REDUCE THE POWER OF PRIVATE PLATFORMS BY:

Building citizen-consumer activism and creating a “sea change in public opinion”.

Scholars and theorists suggest that a shift in public attitudes is needed to persuade digital media companies to change, there is however no empirical data to draw upon as to how effective this approach would or would not be.

COMBAT FAKE NEWS BY:

Developing and circulating persuasive counter-narratives. The focus would need to be on emotional not rational, appeal. This is proposed but unmeasured.

OVERCOME FILTER BUBBLES/ECHO CHAMBERS, AND SILENCING EFFECTS OF HATE SPEECH BY::

Supporting new platform designs with different design affordances.

The design of platform affordance has an impact on inclusion and participation, as well as the types of interactions people experience and information they are exposed to. There is some suggestion that design affordances can reduce the effects of filter bubbles by engaging internet users in more ideologically diverse communities. Well-designed, collectively-owned, online deliberative fora like Loomio have been empirically shown to also create a safe environment for marginalised groups. Research suggests that intentionally building more participatory forms of engagement into platforms might reduce filter bubbles, echo chambers and incivility, while increasing communication and deliberative processes.

IMPROVE TRUST IN DEMOCRACY BY:

The creation, selection and use of online platforms that afford citizen participation and deliberation. Some empirical evidence shows that direct and participatory democratic engagement/processes, e-government, and open government improve trust.

International research has found that engaging citizens in deliberative processes often results in profound changes in deliberating citizens’ “frequently in the direction of more common good-oriented policies.”

Using digital government processes. Transparent, easy to access and well designed e-government and open government initiatives have been shown to increase positive feelings and citizen trust in local government.

CREATE LONG LASTING PROTECTIONS FOR PEOPLE OR INTERVENE TO PROTECT THEM

REDUCE THE POWER OF PRIVATE PLATFORMS, COMBAT INCIVILITY AND MISINFORMATION ONLINE BY:

Improving Content Moderation. Calls for new regulatory policies around content moderation acknowledge this remains an opaque and difficult practice, and on its own is not a fix-all solution. Current policies at the largest intermediaries attempt to balance stakeholder expectations (including users, consumers, advertisers, shareholders, the general public), commercial business goals, and jurisdictional norms and legal demands (which are generally governed by liberal-democratic (US) notions of “free speech”). Goals related to inclusive and participatory democracy are not included.

The most common ‘workable solution’ presented as it relates to content moderation, are processes that combine technical and social (human) responses. Advances in semi or fully automated systems, including deep learning, show increased promise in identifying inappropriate content and drastically reducing the number of messages human moderators then need to review. In the literature however, researchers note that neither automated nor manual classifications systems can ever be “neutral” or free from human bias. Human and/or automated content moderation is unlikely to achieve “civil discourse” or goals through moderation alone. Therefore, the combination of automated classification and deletion systems and human efforts remains the most effective content moderation strategy currently on offer. In the few places where they exist government regulations on private intermediaries’ moderation practices have not been empirically tested for their efficacy or effectiveness.

COMBAT FAKE NEWS BY:

A multi-stakeholder content moderation. This is an approach that combines human and technical intervention, however this is a proposed but untested solution.

REDUCE HATE SPEECH/TROLLING BY:

Using identity verification systems. Sites that do not allow anonymisation and force pre-registration have been shown to solicit qualitatively better, but quantitatively fewer, user comments because of the extra effort required for engaging in discussion. Empirical research has also found that abusive comments are minimised when anonymous commenting is prohibited.

BUILD UNDERSTANDING AND CHANGE INDIVIDUAL BEHAVIOURS

ADDRESS SURVEILLANCE AND DATA PRIVACY ISSUES BY:

Encouraging individuals to employ technical solutions. Such solutions include ad-blockers and ad-tracking browser extensions, private browser options (e.g. Tor), open source platforms and cooperative platform models. “Evidence” supporting the efficacy of these tools and alternatives, however, is typically anecdotal.

COMBATE FAKE NEWS BY:

Education, particularly around critical thinking. Evidence has emerged in health for this approach.

REDUCE HATE SPEECH BY:

Building Resilience through Support Networks. Developing fast and effective reporting mechanisms and support networks. A networked approach can effectively combat the effects of hate speech; by building counter-narratives that counteract racism for example.

Coordinating diverse stakeholders to apply pressure to private intermediaries, in ‘long-haul’ campaigns, has also been effective in having hateful content removed from social media. Speed of removal is considered essential to diffusing the power of hate speech and trolling.

IMPROVE TRUST IN DEMOCRACY BY:

Civics education. Educating children in schools on ‘good citizenship’ has been positively associated with increased political engagement.

CONCLUSIONS

THE THREE CORE PROBLEMS TO EMERGE FROM OUR RESEARCH

At the heart of the challenges to democracy posed by digital media are three core problems:

- Platform monopolies: two or three corporations control not only our means of communication, but also the content which is distributed both of which are core aspects of our democracy, whilst the market power and global mobility of these companies make it possible for them to avoid national regulatory measures either by moving operations elsewhere or simply ignoring them;
- Algorithmic opacity: algorithmic engines are using huge quantities of personal data to make ever more precise predictions about what we want to see and hear, and having ever increasing influence over what we think and do, with little transparency about how they work or accountability for their impact; and
- Attention economy: the dominant business model of digital media prioritises the amplification of whatever content is best at grabbing our attention, while avoiding responsibility for the impact that content has on our collective wellbeing and our democracy. The negative impact is brutally clear from both the literature and the world around us.

KEY PRINCIPLES FOR POLICY RESPONSE:

Use democratic processes, which provide some degree of transparency about the decisions being made, accountability as to their impacts, and opportunities for challenge and judicial review. These processes must include meaningful participation by diverse representatives of the people whose lives are impacted by digital media. In particular, Internet users and civil society must have meaningful involvement, as the crucial third party in the multi-stakeholder process.

Draw on the evidence as to what is most likely to work, where it exists. Perhaps the most predictable finding of this research is that there has been little or no investment by people in government or other research funders into experimenting with and recording possible solutions, and there needs to be more.

Evidence-led and principled approach. Where there are gaps in the evidence, there are key principles that can be followed to reduce the risk of implementing solutions that do more harm than good. These include an evidence-led focus on ‘upstream’ structural change and the application of human rights principles.

Focus on structural or 'upstream' change. Tackle the structural drivers that underlie all the downstream problems - such as online abuse, disinformation, radicalisation and polarisation. Solutions should be designed to intervene at the structural level and to rebalance power through, for example: governance structures, regulation to restore transparency, accountability and fair competition and genuinely participatory and representative multi-stakeholder processes. None of this is to say that design solutions and platform affordances are not important. As the research shows, they will be essential. But without some rebalancing of power, without increasing the diversity of people involved in decision-making at the highest levels, those design solutions run the risk of replicating very similar problems to those we now face.

Respect and protect human rights. The following human rights principles should also be applied to policy development in this area:

- Universality: Human rights must be afforded to everyone, without exception.
- Indivisibility: Human rights are indivisible and interdependent.
- Participation: People have a right to participate in how decisions are made regarding protection of their rights.
- Accountability: Governments must create mechanisms of accountability for the enforcement of rights.
- Transparency: Transparency means governments must be open about all information and decision-making processes related to rights.
- Non-Discrimination: Human rights must be guaranteed without discrimination of any kind.

Agile approach. In the absence of a strong evidence base, it makes sense to take an agile, iterative approach to policy change. Experiment with all the policies all the time. Ensure that the funding, design, and implementation of policies reflect a record, learn, and adapt approach to measure the impact of any new initiatives or regulations, and to make adjustments as evidence becomes available as to impact.

URGENT AREAS FOR CHANGE

Some of the areas in which action is needed sooner rather than later include effort to:

Restore a genuinely multi-stakeholder approach to internet governance, including rebalancing power through meaningful mechanisms for collective engagement by citizens/users;

Refresh antitrust & competition regulation, taxation regimes and related enforcement mechanisms to align them across like-minded liberal democracies and restore competitive fairness, with a particular focus on public interest media;

Recommit to publicly funded democratic infrastructure including public interest media and the creation, selection and use of online platforms that afford citizen participation and deliberation;

Regulate for greater transparency and accountability from the platforms including algorithmic transparency and great accountability for verifying the sources of political advertising;

Revisit regulation of privacy and data protection to better protect indigenous rights to data sovereignty and redress the failures of a consent-based approach to data management; and

Recalibrate policies and protections to address not only individual rights and privacy but also to collective impact wellbeing. Policies designed to protect people online need to have indigenous thinking at their centre and should also ensure that all public agencies responsible for protecting democracy and human rights online reflect, in their leadership and approaches, the increasing diversity of our country.

ACKNOWLEDGEMENTS

This research was made possible by support from our major funder, the New Zealand Law Foundation's Information Law & Policy Project, with additional research funding from the Luminate Group.

The research team on this project was lead by Marianne Elliott (The Workshop) and included Dr Jess Berentson-Shaw (The Workshop), Dr Kathleen Kuehn (Victoria University of Wellington), Dr Leon Salter (Massey University) and Ella Brownlie (The Workshop). Project management was provided by Jay Brooker (The Workshop). The quantitative survey was conducted by UMR Market Research.

PURPOSE, DEFINITIONS AND METHOD

RESEARCH PURPOSE

Digital media has been heralded as inherently democratising. People have direct access to each other and to their elected representatives, across geographical and cultural boundaries. But increasingly it is also seen as a space in which democracy may be simultaneously undermined.

As digital technology increasingly permeates society, there is good reason to pay attention to the institutions, policies, and practices that surround this technology and present both opportunities and threats to democracy. This is especially true for government, as those people who represent the interests of all citizens, but it is also true for everyone with an interest in the future health of our democracy.

The purpose of this research was to explore the opportunities, risks and threats posed to New Zealand's democracy by digital media, in order to scope future research into the policy solutions available to New Zealand to maximise the opportunities, and to meet and mitigate the threats.

DEFINITIONS

In order to assess the impact of something like digital media on democracy, you need a definition of democracy. We used a definition of democracy adapted from the framework developed by the Economist Intelligence Unit for their Democratic Index report, and the definition used by Jamie Bartlett in his book 'The People vs Tech'.

The five features of democracy in our definition are:

Electoral process and pluralism: including whether elections are free, fair and trusted.

Active citizens: alert, informed citizens who are capable of making important moral judgements, including measures of equity and diversity in representation.

Shared democratic culture: enough societal consensus, cohesion and willingness to compromise for a stable, functioning democracy. In New Zealand, this includes

compliance with te Tiriti o Waitangi, on which our democratic culture is founded.

Civil liberties and competitive economy: a functioning competitive economy and civil society, including protection of human rights and free, independent media.

Trust in authority: a trustworthy government, parliament and judiciary and elected representatives accountable to the people.

Defining democracy is complex. Defining digital media is almost as difficult. Digital media technically includes all digitised content that can be transmitted over the internet or computer networks. This could include text, audio, video, and images. So content from print or broadcast media outlets can fall into this category when it is presented on a website or blog.

The focus of this research was on social media, online forms of communication that people use to share and exchange information with interested audiences, and within that, a specific focus on the major digital platforms, including Facebook, YouTube and Twitter. However, interviewees also talked about the impact of other forms of digital media on democracy, including blogs, online forums and digital forms of traditional media.

METHODOLOGY

This research project was made up of three separate but related strands: expert interviews, a literature review, and a quantitative survey.

EXPERT INTERVIEWS

A series of in-depth interviews were conducted with experts and stakeholders to explore the scope of this issue in more detail, prioritise various aspects of the problem for future research and identify key potential collaborators for further research.

Thirty-five in-depth interviews were undertaken with a selection of experts, stakeholders and users drawn from the following sectors.

- Political
- Policy and official
- Māori-led organisations
- Civil society
- Industry/sector organisations
- Academics, researchers and experts
- International experts.

The interviews were recorded and transcribed and analysed using a hybrid of content and grounded analysis in which some broad themes were used as a starting framework for the analysis, but amended and altered based on the themes that emerged from the data as the analysis progressed.

The full report on the interviews can be accessed [here](#)

LITERATURE REVIEW

The literature was conducted in two parts, one looking at the nature of the opportunities and threats to democracy from digital media, and the other looking at the evidence as to effective solutions and responses.

Part one: Opportunities and Threats

In this narrative literature review, we sought to describe, from the most recent literature (searches were limited to research published in the last eight years and most are within five), what the nature of the opportunities and threats are to democracy from developments in digital media.

We asked two research questions:

1. What are the specific opportunities digital media presents for improving democratic participation?
2. What are the current threats/barriers that are in place to prevent achieving those opportunities?
3. A non-systematic narrative review was chosen with a view to summarising the themes that have been covered in terms of opportunities and problems (risks and threats). Searches were limited to research published in the last eight years (most are within five).
4. In total, 110 documents were reviewed including journal articles, reports and book chapters.

Part Two: Solutions

Following on from the review of the literature identifying the opportunities and threats that digital media pose to an inclusive and participatory democracy (Part One), we undertook a review to identify tested and workable solutions to realising the potential of digital media and/or overcoming current threats.

A non-systematic narrative review was chosen with a view to summarising the evidence. Searches were limited to research published in the last eight years (most are within five). It was not an exhaustive review, but in general we found a dearth of empirically tested solutions. This is not surprising given the relatively slow response of government and other public institutions (from where such research would most logically be situated and or funded) to the threats from digital media.

The review is presented in three parts: 1) the empirical evidence on workable solutions to threats to democracy from digital media, 2) a summary of recommendations found in the literature and 3) a brief discussion of some activities identified in New Zealand

The full literature reviews can be accessed [here](#).

QUANTITATIVE SURVEY

The third component of this research was an online survey among a nationally representative sample of New Zealanders aged 18 years and over. The survey was designed to elicit the views and experiences of people using social media and digital platforms relevant to democracy (e.g. participating in debates about issues of public policy on social media.)

1,000 people were surveyed, weighted to accurately reflect the New Zealand population in relation to region, age, gender and ethnicity. Fieldwork was carried out from the 27th of September to 2nd of October 2018.

The full report on the survey can be accessed [here](#).

OVERVIEW

Over recent years a growing body of international research has looked at the impact of digital media on democracy, with particular focus on the US and the UK where the role played by digital media in the election of Trump and the Brexit referendum raised significant concerns.

This project was designed to find out if we should be worried about these same issues here in New Zealand. And if so, what should we do about it? In order to answer that question we identified five key features of democracy against which we could measure the impact of digital media, for better and for worse. They are:

- Electoral process and pluralism
- Active, informed citizens
- Shared democratic culture
- Civil liberties and competitive economy
- Trust in authority

WHAT WE'VE FOUND

Critically, we found that digital media is having an impact across every one of those features of a healthy democracy.

There are indicators that digital media has had some beneficial impacts. Our quantitative research here in New Zealand indicates, for example, that people from minority groups have been able to use digital media to participate in democratic processes including accessing politicians and engaging in public debate. Whatever our response to the challenges posed to democracy by digital media, it's important we don't lose these opportunities in the process.

But the overall trend should raise serious concerns. Active citizenship is being undermined in a variety of ways. Online abuse, harassment and hate - particularly of women, people of colour, queer people, people with disabilities and people from minority religions - undermines democratic participation not only online, but offline.

Misinformation, disinformation and mal-information are undermining not only informed debate, but also public trust in all forms of information. Distraction and information overload are eroding citizens' capacity to focus on important and complex issues, and their capacity to make the 'important moral judgements' required in a functioning democracy.

Likewise, interviewees described a myriad of ways in which our shared democratic culture is being undermined by digital media - including through disinformation, polarisation, attention hijacking and radicalisation.

One of the clearest impacts of digital media on our democracy has been its impact on funding for mainstream media. While Facebook and Google Hoover up the advertising revenue that once would have been spent on print, radio and television advertising, they contribute nothing to the work of producing the kind of news and current affairs reporting that is essential to a functioning democracy. In a stunning display of hypocrisy, Facebook recently complained that their local news service was being hindered by a lack of local newspapers, many of which were forced to either shut down or significantly reduce their newsroom size after losing advertising income to Facebook.

The representative survey we carried out indicates that New Zealand's small size and relatively healthy mainstream media (relative to elsewhere and despite significant resource challenges) may help us avoid the worst effects of "filter bubbles" and "echo chambers" in digital media on some issues.

When asked about the legalisation of cannabis, New Zealanders who got their information about the issue online were able to predict relatively accurately whether the majority of New Zealanders shared their views or not. A third of those who disagreed could predict (that is a minority), most who agreed could accurately predict. This may be unique to the debate about drug reform because, for example, there had been significant media coverage of opinion polls on this issue. More research would be needed to see if this is replicated across other issues in New Zealand.

Interviewees in our qualitative research nonetheless pointed to examples where debate in New Zealand about issues like free speech, hate speech and gender identity attracted the attention of foreign actors holding strong, even extreme, views on these issues. Engagement by these foreign actors in the online public debates on issues here in New Zealand appears to some interviewees to have contributed to a polarisation, even radicalisation of views here. Interviewees also raised concerns that the ability of citizens to form free and informed opinions were being undermined not only by mis and disinformation, but by the increasing role of algorithms in predicting and curating the information each of us is exposed to.

THE NEED FOR A SYSTEMIC RESPONSE

We could continue to outline the impact digital media is having on trust in public institutions, free and fair elections, the protection of human rights and a competitive economy. More on all of that below. The key message is clear, digital media is having massive, system-wide impacts on our democracy. It affects every part of our lives and the people who run the corporations controlling the major platforms are having a determinative impact on the very structures and functions of our society. While better content moderation is clearly one of the responses we must demand of the platforms, it is not even close to being a sufficient response to the scale of the challenge.

THE THREE CORE PROBLEMS TO EMERGE FROM OUR RESEARCH

At the heart of the challenges to democracy posed by digital media are three core problems:

5. **Platform monopolies:** two or three corporations control not only our means of communication, but also the content which is distributed both of which are core aspects of our democracy, whilst the market power and global mobility of these companies make it possible for them to avoid national regulatory measures either by moving operations elsewhere or simply ignoring them;
6. **Algorithmic opacity:** algorithmic engines are using huge quantities of personal data to make ever more precise predictions about what we want to see and hear, and having ever increasing influence over what we think and do, with little transparency about how they work or accountability for their impact; and
7. **Attention economy:** the dominant business model of digital media prioritises the amplification of whatever content is best at grabbing our attention, while avoiding responsibility for the impact that content has on our collective wellbeing and our democracy. And the negative impact is brutally clear from both the literature and the world around us.

It's critical that this moment of global cooperation is used to address the wider, structural drivers of the biggest threats posed to democracy by digital media. These structural drivers include the power that a handful of privately-owned platforms wield over so many aspects of our lives, from what information we see, who we interact with, and who can access information about us. And we must do this while maintaining and building upon the many opportunities digital media simultaneously offer to tackle some of the biggest challenges facing democracy, including inequity of access and declining engagement.

In order to do that, action is needed sooner rather than later in order to:

- **Restore** a genuinely multi-stakeholder approach to internet governance, including rebalancing power through meaningful mechanisms for collective engagement by citizens/users;
- **Refresh** antitrust & competition regulation, taxation regimes and related enforcement mechanisms to align them across like-minded liberal democracies and restore competitive fairness, with a particular focus on public interest media;
- **Recommit** to publicly funded democratic infrastructure including public interest media and the creation, selection and use of online platforms that afford citizen participation and deliberation;
- **Regulate** for greater transparency and accountability from the platforms including algorithmic transparency and accountability for verifying the sources of political advertising;
- **Revisit** regulation of privacy and data protection to better protect indigenous rights to data sovereignty and redress the failures of a consent-based approach to data management; and
- **Recalibrate** policies and protections to address not only individual rights and privacy but also collective dynamics and wellbeing, and protect indigenous rights. Public agencies responsible for protecting democracy and human rights online should reflect, in their leadership and approaches, the increasing diversity of our country.

THE OPPORTUNITIES OF DIGITAL MEDIA

The potential of digital democracy lies in its ability to increase democratic participation, embrace diversity of opinion, and empower marginalised groups. We identified six clear opportunities from the literature that digital media offers. These are: the democratisation of information publishing, broadening the public sphere, increasing equality of access to and participation within political processes, increasing participation and engagement in political processes, increasing transparency and accountability from government and promotion of democratic values.

Broadly speaking these opportunities fit into two categories: First, those that enable individuals, citizens or groups, who due to their status in society have been excluded from fully participating in different aspects of the democratic process, through greater access to the levers of democracy. Examples include the use of digital media to: broaden the public's engagement with indigenous people and their lives, give more exposure to women in politics, build well-networked, educated and empowered communities, and encourage political engagement from youth.

The second category of opportunities relate to digital media's use by people in governments to make the processes of democracy more inclusive, to increase engagement with citizens, improve transparency of government work, and rebuild trust in democratic processes. Examples of such work include online deliberative democracy processes, open or e-government initiatives, and funding of public service journalism, platformed on digital media.

DEMOCRATISATION OF INFORMATION PUBLISHING

Digital media enables the creation and sharing of content by anyone. This aspect in particular, the literature shows, has the potential to improve democratic participation by facilitating dialogue both between governments and citizens (improving institutional trust) and between otherwise divergent groups and individuals in society.

BROADENING OF THE PUBLIC SPHERE

The literature suggests that digital media can be used to widen policy conversations to include marginalised individuals and communities who have been previously excluded from democratic processes. A good example of this is the @IndigenousX Twitter account.

INCREASING EQUALITY OF ACCESS TO AND PARTICIPATION WITHIN POLITICAL PROCESSES

Several studies found that digital media increases equality of access to and participation within political processes, in terms of gender, class, race and age. Specifically digital and social media:

- gives more positive exposure to women politicians than traditional news media;
 - builds well-networked, educated and empowered communities, which may previously have been economically and socially marginalised by digital divides, (when incorporated with other good government policies such as civics education),
 - facilitates the formation of both 'ad-hoc' and longer-term, group-based online communities focused on fighting racism, which can provide a safe space of belonging for ethnic minority groups.
 - softens political inequality patterns by encouraging political engagement from 16-29 year olds.
-

INCREASING ENGAGEMENT IN POLITICAL PROCESSES

Numerous studies found links between digital media and increased engagement in political processes in the general population, not only in marginalised groups. This includes engagement in elections, different forms of deliberative democracy, as well as participation in more informal political action such as protests.

INCREASING TRANSPARENCY AND ACCOUNTABILITY FROM GOVERNMENT

While digital media has the potential to help rebuild trust in public trust in democratic institutions and policies using “open government” and “e-government” initiatives, the research suggests it is likely dependent on the base level of trust (i.e such approaches may be more effective in low trust environments where there is less initial transparency).

PROMOTION OF DEMOCRATIC VALUES

People in governments have been able to actively promote democratic values, informed debate, tolerance and respect for other groups using digital media. Examples of direct action include government funding of public service journalism, funding of independent statutory organisations such as All Together Now in Australia, which encourages embracing of cultural diversity. Less direct action includes the use of digital media to promote participatory democracy activities, e.g deliberative forums.

SUMMARY

The opportunities for digital media are significant and important. If used well, digital media can enable governments to respond effectively to the experiences of marginalised groups, to ensure equitable policies and practices are designed, delivered and adjusted, and to build trust in the democratic institution as responding to the needs of all people. It offers as much to people pushing against barriers to their progression, inclusion, and improved wellbeing in society as it does to people in government looking to remove those barriers and build a more inclusive democratic system.

THE THREATS

The threats to this promise outlined in the literature are significant however, and most are intricately bound up with the concentration of power in profit-driven companies. The seven key threats we identified to inclusive democracy from digital media were: the increasing power of private platforms, foreign government interference in democratic processes, surveillance and data protection issues, fake news, misinformation and disinformation, filter bubbles and echo chambers, hate speech and trolling, and distrust/dissatisfaction with democracy.

Some of these threats or problems originate in the structures and systems of society e.g the power of private platforms over people's lives. Others operate at an individual level e.g. a growing distrust of democracy. However, all these threats are interconnected. Together they threaten to derail the democratic promise of digital media.

INCREASING POWER OF PRIVATE PLATFORMS

Private platforms have increasing power to determine all aspects of our access to information, social interactions, and democratic activities. Researchers highlighted the increasing dominance of an increasingly small number of privately-owned platforms over the internet. People who own and control these platforms have a monopolisation tendency linked to the relationship between the mining of user-data and their imperative to make profits. This model of operation is termed “platform monopolies”. The monopolisation tendency makes alternatives to the data-extraction for profit model, for example co-operative, democratised ownership models, hard to start up and survive.

The concentrated power of these platforms shapes not just the wider information context and ability to develop alternative non-extractive models of digital information provision and sharing, but individual’s personal lives also. Platform monopolies affect how and with whom we interact socially through algorithms. A body of literature points to the actions that these people in this companies take that impact human rights, both through the control of personal data and the level of control over what appears in the public sphere.

From this model of platform monopolies flows a series of further threats to democracy. Some relate to the features of the platforms, directly linked to the capture of people’s personal data. The collection and on-sale of personal data by these platforms, to both governments looking to undertake surveillance on their own citizens, and private organisations wishing to make profits, erodes public trust in information systems, and curtails the professional work of the media and writers - a key plank in our democracy.

FOREIGN GOVERNMENT INTERFERENCE IN DEMOCRATIC PROCESSES

The literature shows that the interference in democracy, specifically through the use and manipulation of digital and social media contributes to decreased turnout and voter disengagement. Disinformation campaigns by foreign governments exaggerate already existing tensions and polarisations and encourage a lack of faith in the electoral system and lack of trust in the idea of liberal democracy.

SURVEILLANCE AND DATA CAPTURE

Evidence shows that the provision of private data to both governments for the purposes of surveillance and private organisations for profit-making by platforms has a curtailing effect on key components of democratic function. Specifically, there has been a demonstrated curtailing effect on the private and public practices of both writers and journalists. While surveillance of Muslim communities for example contributes to alienation from mainstream society.

FAKE NEWS, MISINFORMATION AND DISINFORMATION

The creation of the “attention economy” also poses a significant threat. People’s propensity to attend to shocking, false, or emotive information, especially political information, is exploited and used as a commodity product by digital media platforms.

The literature shows that governments with the means and inclination to manipulate information can tailor false information towards individuals with the express intent of interfering in other countries democratic processes, for example Russian government interference in the U.S election of 2016 using ‘bots’ and disinformation campaigns. While misinformation and disinformation, especially political disinformation, targeted at individuals on digital media, is used to influence politics, from national elections through to information provision and sharing with regard to political issues and policy more generally. Political misinformation in particular has been found to have a significant direct and indirect impact on democratic participation and engagement.

FILTER BUBBLES AND ECHO CHAMBERS

Filter bubbles are specific technical effect of the attention economy. Facebook's news feed is a filter bubble, created by a machine-learning algorithm which draws on data created by user networks, likes and comments and how much organisations are willing to pay to be present there.

Filter bubbles follow a longer-term trajectory within advertising (including political advertising and now disinformation) which has sought to collect data in order to tailor adverts to target groups, however, now they can be targeted to specific individuals. This can contribute to the formation of echo chambers, which is the reinforcement of existing beliefs (confirmation bias) through selective exposure to information. Hence, the technical and economic drivers of filter bubbles can act to reinforce echo chambers.

Increasing numbers of automated social media 'bots' have been linked with the spread of political disinformation and thus the reinforcement of echo chambers. Filter bubbles and the related echo chambers they feed into are linked to a decline in trust in the ability to traditional news media to provide reliable information. They have been found to exacerbate political divisions and polarisation, and have negative implications for the mechanisms of liberal democracy. Developing a broad consensus around decisions made in the public good becomes increasingly difficult.

HATE SPEECH AND TROLLING

The rise of hate speech and trolling is linked to polarising effects of filter bubbles and echo chambers. A troll is an anonymous user who deliberately provokes antagonistic reactions for sheer enjoyment. Trolling is aided both by the ease of creating anonymous online profiles and by the atomised nature of internet interaction. Trolling can pose a direct threat to opportunities when it becomes systematically targeted towards minority groups in order to deliberately cause emotional distress. Remaining anonymous makes individuals more likely to escape prosecution for the more egregious examples

Racialised hate speech (otherwise known as cyber racism) is specifically targeted towards ethnic minority groups, and has become increasingly coordinated in recent years, through the rise of the "alt-right". It encourages affected groups to retreat to safe locations, rather than engaging with national debates and institutions.

Sexualised hate speech is primarily targeted towards women (together with members of the LGBTIQ community), and is characterised by its specifically misogynistic nature. It is often directed towards women in the public eye, or those in influential positions, such as journalists, with proponents directing critical attention onto their supposed essential gender characteristics, rather than their work. It has a negative impact on efforts towards the broadening of the public sphere, as women are discouraged from writing what they may feel are controversial stories.

More generally, research has found a correlation between strong, vocal disagreements with an individual's perspectives and a "spiral of silence" which acts to curtail the voicing of contentious opinions by minority groups. The particular ability of trolls and hate speech to fan antagonistic "flames" rather than promote rational debates, has a direct impact on democratic participation.

DISTRUST/ DISSATISFACTION WITH DEMOCRACY

While distrusts with democratic process is a longer-term issue, digital media has likely exacerbated this pattern across western democracies. Researchers argue that trust, informed dialogue, mutual consent, and participation- fundamental features of democracy are being eroded by the features that make social media so profitable. Researchers also found that the way in which the information is distributed on digital media (horizontal, and decentralised and interactive) increases intolerance of others, polarisation and scepticism toward democracy

SUMMARY

The opportunities of digital media, while still apparent, appear to have been suppressed by the sheer weight of fake news, filter bubbles, populism, polarisation, hate speech, trolls and bots, that have emerged from the concentration of power in a small group of private organisations seeking to maximise profits. Digital platforms initially celebrated for their democratic possibilities have transformed into anti-democratic power centres through the collection and exploitation of users attention and data. These privately owned platforms have largely escaped public oversight or regulation over their ability to harness this new power for commercial or political gain.

The question is what can policymakers do to re-calibrate? Are there empirically tested public policies and approaches that can ensure digital media works to strengthen and deepen democracy?

THE SOLUTIONS

A HIERARCHY FOR SOLUTIONS AND INTERVENTIONS

At The Workshop, we take an evidence-informed, hierarchical approach to exploring and understanding problems, and investigating and analysing solutions, policies and practices to overcome them. We work especially to highlight the critical role of structures and systems in improving people's lives with the least individual effort required (though not the least political effort).

IDENTIFYING DRIVERS OF THE PROBLEM

First we ask is the problem we have identified a structural or systems level problem (e.g. the structure of the economic model, the power of private markets over people's wellbeing) or a group/individual level response to the issue (e.g. distrust in democracy that results from a lack of inclusion in democracy)? Sometimes defining where problems originate is complex as there are interactions and feedback loops, as with all complex issues.

For example, hate speech is an individual or group behaviour, it is fundamentally about how people or institutions treat others, however the upstream issues that encourage and enable hate speech, intolerance and bigotry must be explored. Wealth, gender and ethnic inequalities in society, for example, mean digital platforms are primarily owned, designed and managed by those with little experience of differential or harmful treatment based on their position in society.

Without knowledge of how power imbalances and differential treatment based on gender or race play out in society, or a commitment to overcoming them, people who control these platforms can design in policies and practices that encourage hate speech and trolling. By presenting problems in a hierarchy we endeavour to make the feedback loops and upstream structures and systems issues clearer to people.

IDENTIFYING WHERE PEOPLE SHOULD INTERVENE FOR GREATEST IMPACT

In terms of considering "what works", we focus on 'upstream' or structural and systems responses and solutions to the problems. We take this approach because research from across disciplines focussed on enhancing population wellbeing and equity shows interventions at this level:

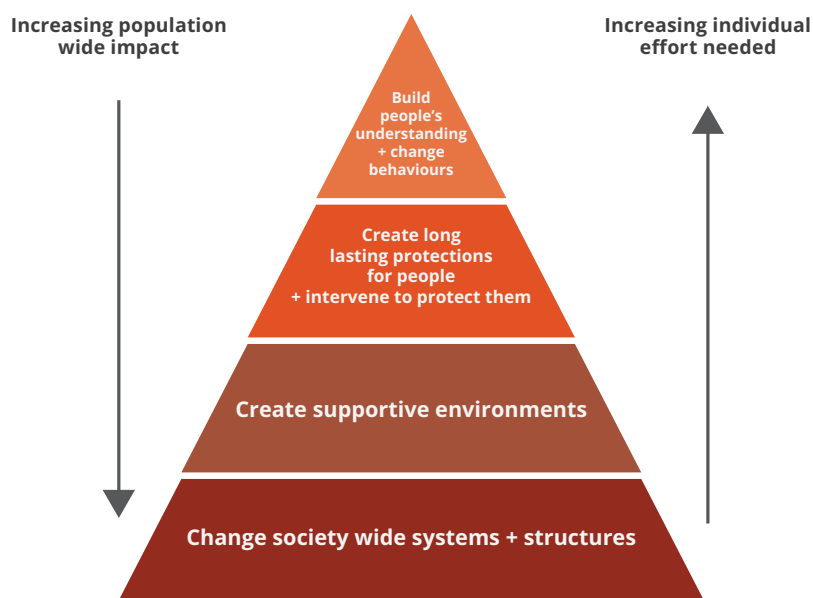
- have the most significant impact on most people's lives and outcomes, and
- require the least effort from individuals to achieve change, and the least resources from those trying to implement change.

We place less emphasis on individual behavioural solutions, not because they are not effective, but because to be effective these solutions (e.g civics education, or consumers closing their Facebook accounts) take significant effort from both individuals and those encouraging such action, and may not address the structural drivers that cause the problems upstream. In addition, people expending energy on individual level solutions can divert energy from investing in understanding and acting on structural level solutions.

The possible interventions identified both in the literature review and by the interviewees are discussed in the context of a hierarchy from those likely to be most effective and requiring least individual effort, to those likely to have the least impact and requiring most individual effort on a population-wide scale. This hierarchy comprises interventions that:

- Change society-wide structural & systems issues to re-establish citizen power
- Create supportive environments & contexts - making the default digital space inclusive and safe
- Create long-lasting protections for people, and intervene to protect them from digital threats
- Build understanding of digital media threats and change individual behaviours in response.

Wellbeing Impact Pyramid



Adapted by The Workshop from Frieden (2010)

FINDINGS

While our literature review was not exhaustive, in general we found a dearth of empirically tested solutions. Likewise, and possibly because of the lack of evidence in the literature, the experts interviewed for this research generally had more to say about the risks and threats they saw arising from digital media than they did on potential workable solutions. However, we did find some common ground between the literature and in the interviews, in terms of solutions.

In line with The Workshop's evidence-led approach set out above, we discuss what empirical evidence we did find in a hierarchy, starting with those solutions likely to have the greatest impact and require least individual effort.

CHANGE SOCIETY- WIDE STRUCTURAL & SYSTEMS ISSUES TO RE-ESTABLISH CITIZEN POWER.

This section focuses on structural and systemic change, addressing for example the disproportionate power of the tech giants vis-a-vis governments, citizens and their domestic competitors.

REDUCE THE POWER OF PRIVATE PLATFORMS BY:

Regulating platforms like other industries. Currently, regulatory debates largely centre around defining the structure, terms and conditions of what kind of industry private intermediaries represent. How platforms should be regulated or governed thus partly hinges on how these services are defined; for example, whether social media platforms are media companies, public spaces, utilities or some other service largely informs how they can ultimately be governed. There is little or no empirical evidence to show how regulation in this area would or would not work, and therefore adaptive approaches to policy and regulation will be needed. This will involve ensuring that the impacts of any change are regularly monitored and changes made as needed in response to those findings.

Introducing new modes of collective action. Under industrial capitalism we had collective bargaining, the strike – e.g., forms of collective action that were sanctioned by law and had the support of a society that allowed people to tame capitalism with legal protection. In relation to digital media researchers suggest there are opportunities for more collective action both by tech workers, demanding for example more ethical design in the products they work on, and by digital media users. New forms of collective, collaborative action that connect users/consumers with the market and state to tame and outlaw surveillance capitalism are suggested by multiple researchers, but again there is no empirical testing to yet drawn upon.

COMBAT FAKE NEWS BY:

Supporting a vibrant and diverse media sphere. One that balances strong, independent and adequately resourced public service media with a non-concentrated commercial media sector. Although there is an existing body of research showing the positive impact of a vibrant and healthy public and independent media on democracy, the specific impact of investing in media in the context of digital media is a widely proposed but as yet unmeasured idea.

CREATE SUPPORTIVE ENVIRONMENTS & CONTEXTS - MAKING THE DEFAULT DIGITAL SPACE INCLUSIVE AND SAFE

REDUCE THE POWER OF PRIVATE PLATFORMS BY:

Designing new competitive digital media solutions. Disruptive technology is needed to forge an alternative digital future that in turn, facilitates a more democratic internet. This means the creation of platforms offering a different set of affordances (ie not those driven by platform monopolies). Platform cooperatives like Loomio subscription-based models and pro-privacy and non-commercial alternatives are already in use and show some evidence of effectiveness in the literature.

REDUCE INTERFERENCE FROM FOREIGN GOVERNMENTS AND POWERS BY:

Designing new anti-cybersecurity infrastructure and drawing upon “big datasets” to review and assess electoral policies. The research in this area is also largely normative, and seems to generally prescribe such infrastructure and reviews will reduce threats to elections and other political processes.

ADDRESS SURVEILLANCE & IMPROVE DATA PROTECTION BY:

Regulating companies’ information management practices. Some regulatory measures, like the Singaporean Data Protection Act 2012, work to and have been proven effective in bringing formal charges to data mismanagement and abuse.

Making regulatory changes to data privacy policies. However, there is little evidence to suggest that these changes will reduce surveillance/data collection so much as regulate how that data is stored, accessed and used by data collectors and other third parties

COMBAT FAKE NEWS BY:

Developing and circulating persuasive counter-narratives. The focus would need to be on emotional not rational, appeal. This is proposed but unmeasured.

OVERCOME FILTER BUBBLES/ECHO CHAMBERS BY:

Supporting new platform designs with different design affordances

Design affordances ascribe meaning to how to use the digital media tool, for example Facebook has a “friend” button directing the user towards ways of interacting based on mutual agreement, also a “share” button, while Twitter has a “follow” button, open to all people using the platform, directing or suggesting different ways of interacting. The design of these affordances has an impact on inclusion and participation, as well as the types of interactions people experience and information they are exposed to. There is some suggestion that design affordances can reduce the effects of filter bubbles by engaging internet users in more ideologically diverse communities.

Non-commercial platforms like Loomio, for example, afford different modes of interaction based on the features (e.g., tools, interface) and environment (e.g., deliberative; asynchronous) it makes available outside a commercial space. The platform affords an environment of less performative, and thus more considered, dialogue, discussion and

debate. The relationship between design and civility on these new platforms has been empirically demonstrated, shown a reduced propensity to engage with similar-minded people encouraged by automated filter bubbles and to move deliberation beyond debate to collective agreement.

OVERCOMING SILENCING EFFECTS OF HATE SPEECH BY:

Supporting new platform designs with different design affordances. Well-designed, collectively-owned, online deliberative fora like Loomio have been empirically shown to also create a safe environment for marginalised groups.

Research suggests that intentionally building more participatory forms of engagement into platforms might reduce filter bubbles, echo chambers and incivility (particularly on mobile devices), while increasing communication and deliberative processes. Therefore, the act of consciously designing social platforms to engender pro-social forms of engagement can have a demonstrated impact on civility

IMPROVE TRUST IN DEMOCRACY BY:

The creation, selection and use of online platforms that afford citizen participation and deliberation. Some empirical evidence shows that direct and participatory democratic engagement/processes, e-government, and open government improve trusts. International research has found that engaging citizens in deliberative processes often results in profound changes in deliberating citizens' "frequently in the direction of more common good-oriented policies", but for them to be effective the systems and platforms used in these deliberative processes must also enable these practices to emerge. The techno-social affordances inherent to different online platforms affect and shape the nature of engagement, deliberation and discussions.

Using digital government processes. Transparent, easy to access and well designed e-government and open government initiatives have been shown to increase positive feelings and citizen trust in local government. Some evidence shows governments that have created usable, intelligible websites, and offer non-exclusionary solutions for those lacking computer and internet access or basic digital literacy skills, have been most successful in their e-government initiatives and constituent satisfaction.

CREATE LONG- LASTING PROTECTIONS FOR PEOPLE OR INTERVENE TO PROTECT THEM FROM DIGITAL THREATS

REDUCE THE POWER OF PRIVATE PLATFORMS, COMBAT INCIVILITY AND MISINFORMATION ONLINE BY:

Improving Content Moderation. Calls for new regulatory policies around content moderation at large intermediaries acknowledge content moderation remains an opaque and difficult practice, and on its own is not a fix-all solution. Current policies at the largest intermediaries attempt to balance stakeholder expectations (including users, consumers, advertisers, shareholders, the general public), commercial business goals, and jurisdictional norms and legal demands (which are generally governed by liberal-democratic (US) notions of “free speech”), goals related to inclusive and participatory democracy are not included.

The most common ‘workable solution’ presented as it relates to content moderation are processes that combine technical and social (human) responses. However, advances in semi- or fully automated systems, including deep learning, show increased promise in identifying inappropriate content and drastically reducing the number of messages human moderators then need to review. In the literature however, researchers note that neither automated nor manual classifications systems can ever be “neutral” or free from human bias. Human and/or automated content moderation is unlikely to achieve “civil discourse,” a “sanitised” internet or other speech and engagement goals through moderation alone. Therefore, the combination of automated classification and deletion systems and human efforts remains the most effective content moderation strategy currently on offer. In the few places where they exist government regulations on private intermediaries’ moderation practices have not been empirically tested for their efficacy or effectiveness.

COMBAT FAKE NEWS BY:

A multi-stakeholder content moderation. This is an approach that combines human and technical intervention, however this is a proposed but untested solution.

REDUCE HATE SPEECH/TROLLING BY:

Using identity verification systems. Sites that do not allow anonymisation and force pre-registration have been shown to solicit qualitatively better, but quantitatively fewer, user comments because of the extra effort required for engaging in discussion. Empirical research has also found that abusive comments are minimised when anonymous commenting is prohibited.

BUILD UNDERSTANDING OF DIGITAL MEDIA THREATS AND CHANGE INDIVIDUAL BEHAVIOURS IN RESPONSE.

REDUCE THE POWER OF PRIVATE PLATFORMS BY:

Building citizen-consumer activism and creating a “sea change in public opinion”. Scholars and theorists suggest that a shift in public attitudes is needed to persuade digital media companies to change, there is however no empirical data to draw upon as to how effective this approach would or would not be.

ADDRESS SURVEILLANCE AND DATA PRIVACY ISSUES BY:

Encouraging individuals to employ technical solutions. Such solutions include ad-blockers and ad-tracking browser extensions, private browser options (e.g. Tor), open source platforms and cooperative platform models. “Evidence” supporting the efficacy of these tools and alternatives, however, is typically anecdotal or prescriptive in nature (as opposed to empirical).

COMBAT FAKE NEWS BY:

Education, particularly around critical thinking. Evidence has emerged in health for this approach.

REDUCE HATE SPEECH BY:

Building Resilience through Support Networks. Developing fast and effective reporting mechanisms and support networks, e.g Advocacy and civil society organisations like All Together Now, have demonstrated some success with building online reporting tools that rely on crowdsourcing to identify – in order to remove - racist hate speech online. A networked approach can effectively combat the effects of hate speech; by building counter-narratives that counteract racism for example.

Coordinating diverse stakeholders to apply pressure to private intermediaries, in ‘long-haul’ campaigns, has also been effective in having hateful content removed from social media. Speed of removal is considered essential to diffusing the power of hate speech and trolling. Pressure from researchers and advocacy groups alike have also encouraged some platforms to design more pro-social tools (i.e., affordances) into their systems.

IMPROVE TRUST IN DEMOCRACY BY:

Civics education. Educating children in schools on “good citizenship” has been positively associated with increased political engagement.

WHAT ROLE FOR NZ?

All of this raises the question: what role should New Zealand play in the wider global efforts to respond to the challenges of digital media? Some interviewees argue that New Zealand should follow the lead of bigger, like-minded liberal democracies like the United Kingdom, the European Union and Australia. Others thought New Zealand could, and should be leading on these issues. Some saw specific opportunities for New Zealand to provide leadership in niches, like indigenous data sovereignty.

One thing many people agreed on was a sense of urgency – an urgency which has increased considerably in the months since many of these interviews took place.

As one participant put it, “we’ve got some really resounding early warning signals about how this stuff can be used to erode our democratic institutions, and if we don’t sit up and take notice of it, and don’t provide the necessary technical, social, and regulatory responses, we might wake up and find that we’ve missed the opportunity.”

NEW ZEALAND AS FOLLOWER

One of the common reasons given for taking the path of following the lead of others was New Zealand's size. However another, perhaps more critical, argument was that New Zealand would need a much better system for making policy on these issues before we can be any kind of global leader. Before we can lead, this participant argued, we need to build up our national capacity to understand and deal with these issues, and build up more of an evidence base, we need to be equipped to have a solid base for developing policy ourselves. It's hoped this research can help contribute to that process.

NEW ZEALAND AS LEADER

On the other hand, some interviewees asked why New Zealand should be a 'taker' of policy on these issues, and identified a great opportunity for New Zealand to team up with other like-minded democracies. We are typically at the cutting edge of technology, they argued, so why not take a lead on this. Digital media has brought advantages to New Zealand, they argued, so we want to make sure that we don't lose the upsides of the new digital economy. Playing a leading role in the global response to the threats of digital media can help ensure that we do not.

NEW ZEALAND AS NICHE INFLUENCER

Some interviewees pointed to New Zealand's track record of taking a principled stand on big global issues, giving our nuclear-free policy as an example. One example given as an area in which New Zealand could show leadership is in the development of a tech workers' union. Because New Zealand has comparatively better employment protections than many other places where tech people work, they said, we already have less of the fear of speaking up. We also have a small enough sector where personal relationships can very easily be brought to bear on these situations.

Finally, but perhaps most importantly, some interviewees argued that there was a role for New Zealand to play as a leader on indigenous data sovereignty and issues relating to Māori digital issues. This would first require us to address the significant gaps in our own protection of indigenous rights online. One of the most critical issues is the need to protect indigenous data sovereignty, allowing Māori ownership and control of Māori data.

CONCLUSIONS

At the heart of the challenges to democracy posed by digital media are three core problems:

1. **Platform monopolies:** two or three corporations control not only our means of communication, but also the content which is distributed both of which are core aspects of our democracy, whilst the market power and global mobility of these companies make it possible for them to avoid national regulatory measures either by moving operations elsewhere or simply ignoring them;
2. **Algorithmic opacity:** algorithmic engines are using huge quantities of personal data to make ever more precise predictions about what we want to see and hear, and having ever increasing influence over what we think and do, with little transparency about how they work or accountability for their impact; and
3. **Attention economy:** the dominant business model of digital media prioritises the amplification of whatever content is best at grabbing our attention, while avoiding responsibility for the impact that content has on our collective wellbeing and our democracy. And the negative impact is brutally clear from both the literature and the world around us.

Combined, these problems pose serious threats to our democracy, so it's critical that our responses to them don't further undermine our democratic institutions. The history of digital media has shown that good intentions can, if not informed by the diverse experiences of users and the research evidence, cause more harm.

INCLUSIVE AND TRANSPARENT PROCESSES ARE CRITICAL

Firstly, as Natasha Tusikov and Blayne Haggart have argued, decisions about what kinds of information we have access to should not be made by a handful of American companies. Nor should our government's role in those decisions take place in backroom negotiations. We need to use democratic processes, which provide some degree of transparency about the decisions being made, accountability as to their impacts, and opportunities for challenge and judicial review. These processes must include meaningful participation by diverse representatives of the people whose lives are impacted by digital media.

MORE RESEARCH NEEDED

Secondly, the stakes are high here, so we must draw on the evidence as to what is most likely to work, where it exists. Perhaps the most predictable finding of this research is that there has been little or no investment by people in government or other research funders into experimenting with and recording possible solutions, and there needs to be more. A list of possible areas for further research is included in the full report.

It's not surprising that there is so little experimental evidence as to the effectiveness of various solutions proposed in the normative literature. Change happens very quickly in this area. Until very recently there has been little investment in research from governments, which would be one of the expected sources of funding for such investigations. More is now urgently needed.

EVIDENCE-LED AND PRINCIPLED APPROACH TO URGENT POLICY DEVELOPMENT

Even in the absence of specific evidence as to the effectiveness of different interventions, there are areas in which action is urgently needed. In those cases, there are key principles that can be followed to reduce the risk of implementing solutions that do more harm than good. As a primary principle, we can take an evidence-informed, hierarchical approach to exploring and understanding problems, and investigating and analysing solutions, policies and practices to overcome them. This involves identifying underlying drivers of the problem, and those interventions which are most likely to have the greatest impact.

We set these principles out in more detail in the section on solutions above, but the key point is that we need to focus on tackling the structural drivers that underlie all the more specific problems outlined above - such as online abuse, the spread of disinformation, radicalisation and polarisation, political interference and manipulation or distraction. Solutions should then be designed to intervene at that structural level addressing and rebalancing power through, for example, governance structures, regulation to restore transparency, accountability and fair competition and genuinely participatory and representative multi-stakeholder processes.

None of this is to say that design solutions and platform affordances are not important. As the research shows, they will be essential. But without some rebalancing of power, without increasing the diversity of people involved in decision-making at the highest levels, those design solutions run the risk of replicating very similar problems to those we now face.

APPLY HUMAN RIGHTS PRINCIPLES

Human rights principles should also be applied to policy development in this area, and are particularly useful where there is an absence of specific research evidence. These principles include:

- > **Universality:** Human rights must be afforded to everyone, without exception.
- > **Indivisibility:** Human rights are indivisible and interdependent, which means in order to guarantee civil and political rights, governments must also ensure economic, social and cultural rights (and vice versa).
- > **Participation:** People have a right to participate in how decisions are made regarding protection of their rights. Governments must engage and support the participation of civil society on these issues.
- > **Accountability:** Governments must create mechanisms of accountability for the enforcement of rights. There must be effective measures put in place for accountability if those standards are not met.
- > **Transparency:** Transparency means governments must be open about all information and decision-making processes related to rights. People must be able to understand how major decisions affecting their rights are made and how public institutions responsible for implementing rights are managed.
- > **Non-Discrimination:** Human rights must be guaranteed without discrimination of any kind. This includes not only purposeful discrimination, but also protection from policies and practices which may have a discriminatory effect.

Each of these principles should be applied in the development of a multi-stakeholder response to the threats to democracy posed by digital media.

AGILE AND RESPONSIVE APPROACH TO POLICY

Finally, in the absence of a strong evidence base, it makes sense to take an agile, iterative approach to policy change. Experiment with all the policies all the time. Ensure that the funding, design, and implementation of policies reflect a record, learn, and adapt approach to measure the impact of any new initiatives or regulations, and to make adjustments as evidence becomes available as to impact.

URGENT AREAS FOR ACTION

Some of the areas in which action is needed include efforts to:

- **Restore** a genuinely multi-stakeholder approach to internet governance, including rebalancing power through meaningful mechanisms for collective engagement by citizens/users;
- **Refresh** antitrust & competition regulation, taxation regimes and related enforcement mechanisms to align them across like-minded liberal democracies and restore competitive fairness, with a particular focus on public interest media;
- **Recommit** to publicly funded democratic infrastructure including public interest media and the creation, selection and use of online platforms that afford citizen participation and deliberation.;
- **Regulate** for greater transparency and accountability from the platforms including algorithmic transparency and great accountability for verifying the sources of political advertising;
- **Revisit** regulation of privacy and data protection to better protect indigenous rights to data sovereignty and redress the failures of a consent-based approach to data management; and
- **Recalibrate** policies and protections to address not only individual rights and privacy but also to collective impact wellbeing. Policies designed to protect people online need to have indigenous thinking at their centre and should also ensure that all public agencies responsible for protecting democracy and human rights online reflect, in their leadership and approaches, the increasing diversity of our country.

In the wake of the Christchurch mosque attacks, a new global momentum has emerged around the role that social media has played in the spread of violent extremism and terrorism, and what can be done to stop it. The New Zealand government has, rightly, stepped up to play a leadership role in that work.

What we need right now is a clear analysis of the wider structural and systemic issues that underpin the immediate moderation challenge and a solid proposal of regulatory and other changes that are needed to tackle those bigger issues. That means ensuring that the current, heightened public debate on the role of digital media in fostering and spreading hate is placed into a wider context of the regulatory and structural changes needed to revive and restore the key features of a healthy and peaceful democracy in our country, and around the world.

Our intention is that this research will help frame, inform and support those efforts.

SUMMARY OF KEY FINDINGS FROM EACH PART OF THE RESEARCH

KEY FINDINGS FROM THE SURVEY

Some key findings from the analysis of the survey data:

- > Use of social media is high, and Facebook dominates with Twitter and Instagram also used.
- > Around a quarter of the sample used social media to engage with ‘political issues’ or politicians.
- > Social media platforms are used for political activity by minority ethnic groups more than Pākehā. Indicating their potential use as a tool for engagement in formal democratic system.
- > Stated trust in news online may be low, but perceptions of information credibility are driven by trust in friends, family and organisations.
- > Most people still rely on mainstream media for information about a key political issue (decriminalisation of marijuana) but friends and family and online news feature strongly.

New Zealanders are relatively accurately informed about the opinions of others with regard to the decriminalisation of marijuana.

There is evidence that New Zealanders who believe their views are in the minority on decriminalisation of marijuana are less willing to share their views both offline and online. This suggests social media platforms replicate rather than overcome existing barriers to engaging in less formal processes of democracy (public discussion and discourse) for people who hold minority views.

KEY FINDINGS FROM THE LITERATURE REVIEW

To paraphrase classical historian Mary Beard, western democracy is a 2000 year old experiment. In 2019 the significant technical disruption that is digital media is having a powerful effect on the results. Yet what is the nature of that effect? Does our collective written and published knowledge tell us what benefits and opportunities digital media offers in building a stronger, more inclusive and participatory democracy? And the threats or risks it poses to it? And what if anything does the empirical evidence tell us optimises the opportunities and reduces the risks to our democracy from digital media?

The answer to these questions remains elusive. While our literature review was not exhaustive, this research confirms that there is, at present, a troubling dearth of scientific, empirical, evidence-based research that tests or aims to validate “workable solutions” to the seven key threats to democracy we’ve identified in this project.

While some empirical evidence exists, notably in the area of designing new platforms and affordances with pro-social intent, the significant majority of the research relating to the threats we identified is based on expert opinion and normative approaches. Meaning, it presents theoretically sound arguments about the way things “ought to be” if democracy is to be “reclaimed” from incivility and a rogue form of capitalism in the digital age.

In the expert opinion literature the following four themes were identified:

1. **Policy / Legal Solutions**

For example, adapt existing legislation; create new legislation; institute new oversight bodies or inter-government agencies; or to improve regulations on content moderation.

2. **More Corporate Transparency**

Currently the lack of transparency around moderation practices presents challenges to accountability, governance, and the ability to apply public and legal pressure. Expanding empirical research to improve moderation processes requires private intermediaries to make these processes and practices accessible to researchers.

3. **Better Design**

Platform design can influence the way individuals, organisations and institutions make decisions around platform uses/objectives. Pro-social and democratic values must be encoded into the infrastructure of the internet, including algorithms. At present, the normative values embedded into these global private intermediaries – e.g., openness, connectedness, free speech, etc. – are not culture-neutral norms. It the first step towards designing more deliberative spaces, pro-social tools and online environments.

4. **Improve Content Moderation**

Calls range from the standardisation of industry-wide “best practices” to more transparency and researcher access. These actions would require greater corporate transparency, corporate grievance mechanisms that are transparent, accessible and in accordance with international human rights law, and multi-stakeholder, and inclusive governance approach, and content moderation should become an organisational priority rather than department silo.

This absence of tested solutions is not evidence that proposals do not work, but that they are untested. This leads us to conclude there is a critical need for investment in more research. People in government, civil society, NGOs, and private enterprise need to commit to researchers and projects who will do pre-and post-testing of solutions that stakeholders are recommending.

Such research will not only measure effect and enable us to extend what's working to other places or contexts, but ensure future normative prescriptions are informed by evidence beyond the anecdotal (or budgeting restrictions).

It is critical that people in the New Zealand government especially measure whether or not what is being done is working to build a more inclusive and participatory democracy. New Zealand would break significant ground in that regard.

When people in government and civil society seek recommendations for solutions, they need to mitigate the risk that experts reproduce "solutions" that fit the professional discourses in which they're embedded. To do this, it is important that people in government ask multi-stakeholder group participants:

- > What if any evidence do they have for the suggestions made?
- > What experiences inform these recommendations and why do they identify them as workable solutions over others?
- > How do they imagine testing their effectiveness?
- > Given the current lack of evidence, it is critical that the values, experiences, and outcomes that underlie recommendations are made transparent and visible.

KEY FINDINGS FROM THE INTERVIEWS

One of the questions we posed in the interviews was what had changed in the landscape of democracy, through the influence of digital media, and what has not changed.

A FAMILIAR BUT CHANGING LANDSCAPE

Many things have not changed, participants told us. Misinformation, disinformation and harassment are not new. Outrage, political polarisation and extremism are not new. Filter bubbles, soundbite politics and data capture - none of this is new. Even the erosion of the authority of published material isn't new. And perhaps most tellingly, the cultural hegemony of tech isn't new. Some participants argued that despite all that has changed as a result of digital media, the replication of existing power structures in the governance and management of the tech giants has inevitably lead to reinforcing many already entrenched power imbalances. Further, they said, the lack of diversity at the governance and senior management level prevented these companies from identifying and responded adequately to the risks and threats inherent in their platform designs.

So what has changed? While recognising that the foundations of mis- and disinformation, online harassment and abuse, polarisation and extremism all existed well before the rise of digital media, most interviewees nonetheless saw particular ways in which the

features and functions of digital media has changed the scale, intensity and reach of those phenomena. Digital media has changed the scale, speed and breadth at which information can be shared. It has allowed advertisers, including political advertisers, to target people with much greater precision. Digital media has generated new levels of distraction, undermining citizens capacity to engage in the complex thinking demanded in a democracy. Data has taken on a new value, and has been gathered and used at an unprecedented scale. And finally, but again, perhaps most importantly, a very small number of very large companies control the means of communication used by the majority of people in most democracies on the planet.

So given what has changed with the rise and digital media, what has stayed the same, and the structural underpinnings of the major digital platforms - where are the biggest opportunities for democracy? The obvious and most commonly cited opportunities were in the democratisation of information, increased diversity in public discourse, more public engagement with government and democratic process, and in increased transparency and openness in government.

On the other hand, participants described considerable risks and threats to democracy including digital exclusion. The most commonly cited risks were the impact of digital monopolies, lack of competition and their impact on public-interest media and misinformation and disinformation, including deepfakes and the consequent erosion of trust in information. Other commonly cited risks include political manipulation including foreign interference, cybersecurity of government and security of elections, and the more common manipulation through political advertising, and related risks of polarisation, radicalisation and 'echo chambers'. Other significant risks highlighted by participants were the impact on democracy of online abuse and hate, disengagement, distractions and attention hijacking, and loss of privacy and consent fatigue. Woven throughout many interviews was a recognition that a lack of transparency and accountability by the big platform companies underpinned and exacerbated all of these risks.

As one participant put it, overall, the picture of how democracy as a form is evolving under the influence of digital media is 'quite messy'. "[I]t's got all these new ways to participate, all these new channels for participation. At the same time, it's getting harder to curate and access that content online, and also critique it. So it's a messy space to talk about risks and opportunities, because the whole landscape is so complicated and moving."

SOLUTIONS

Interviewees suggested a range of interventions and solutions to both maximise the opportunities for democracy presented by digital media and minimise the threats. These range from interventions at the structural and systemic level through to suggestions for individual behavioural change.

Some of the areas in which action was identified as being most urgent include effort to:

- **Restore** a genuinely multi-stakeholder approach to internet governance, including rebalancing power through meaningful mechanisms for collective engagement by citizens/users;
- **Refresh** antitrust & competition regulation, taxation regimes and related enforcement mechanisms to align them across like-minded liberal democracies and restore competitive fairness, with a particular focus on public interest media;
- **Recommit** to publicly funded democratic infrastructure including public interest media and the creation, selection and use of online platforms that afford citizen participation and deliberation.;
- **Regulate** for greater transparency and accountability from the platforms including algorithmic transparency and great accountability for verifying the sources of political advertising;
- **Revisit** regulation of privacy and data protection to better protect indigenous rights to data sovereignty and redress the failures of a consent-based approach to data management; and
- **Recalibrate** policies and protections to address not only individual rights and privacy but also to collective dynamics and wellbeing, and protect indigenous rights. Public agencies responsible for protecting democracy and human rights online should reflect, in their leadership and approaches, the increasing diversity of our country.

APPENDIX 1: LITERATURE REVIEW PART 1

DIGITAL THREATS TO DEMOCRACY

> > > > > > > > > > > >

> > > > > > > > > > > >

> > > > > > > > > > > >

> > > > > > > > > > > >

> L I T E R A T U R E > >

> > R E V I E W > > > > >

> > > > P A R T > 1 : > >

> > > T H R E A T S > > >

> > > > A N D > > > > >

O P P O R T U N I T I E S

> > > > > > > > > > > >

> > > > > > > > > > > >

> > > > > > > > > > > >

> > Leon Salter, Kathleen Kuehn, Jess Berentson-Shaw & Marianne Elliott

> > > > >

> >



Luminate

D I G I T A L T H R E A T S T O D E M O C R A C Y

This report is part of the Digital Threats to
Democracy research project.

To see the rest of the reports and the
overall findings go to digitaldemocracy.nz

ISBN: 978-0-473-48026-4

CONTENTS

INTRODUCTION	69
SPECIFIC OPPORTUNITIES FOR IMPROVING DEMOCRATIC PARTICIPATION THROUGH DIGITAL MEDIA	71
2.1 – Democratization of information publishing	72
2.2 – Broadening the public sphere	72
2.3 – Increasing equality of access to and participation within political processes	73
2.4 – Increasing participation and engagement in political processes	74
2.5 – Increasing transparency and accountability from government	75
2.6 – Promotion of democratic values	76
CURRENT THREATS/BARRIERS TO ACHIEVING OPPORTUNITIES	77
3.1 – Increasing power of private platforms	78
3.2 – Foreign government interference in democratic processes	79
3.3 – Surveillance and data protection	79
3.4 – Fake news/disinformation	80
3.5 – Filter bubbles/echo chambers	81
3.6 – Hate speech and trolling	82
3.7 – Distrust/dissatisfaction with democracy	83
CONCLUSION	84
REFERENCES	85

INTRODUCTION

From the early development of digital media, and in the wake of large scale democratic action, including the Occupy movement and the Arab Spring, there was optimistic academic consensus on the capacity of digital media to increase democratic participation. However, the election of Donald Trump and the Brexit referendum, have shaken the foundations of Western democracies, and turned that optimistic view to significant concerns about the role of digital media in eroding democratic participation. There is now clear evidence of interference by the Russian government in the 2016 US presidential election using digital media strategies, which had the effect of discouraging sections of the public from voting (Persily, 2017). Fake news, filter bubbles, populism, polarisation, hate speech, trolls and bots are firmly embedded into mainstream understanding of digital media. All this is in a context where election turnout and trust in government institutions were already in a general decline in Western nations (Leininger, 2015; OECD, 2017).

Rather than the increasingly widespread adoption of digital media necessarily leading to a pattern of increased participation, diversity of opinion, and the empowerment of marginalised groups, digital media (in particular social media through its algorithmically calculated news feeds), can work to create opinion silos, or “echo chambers”, which can “limit the possibility of understanding differences and increase the likelihood of intolerance and hostility” (Lu & Yu, 2018, p. 3).

In this narrative literature review we sought to describe, from the recent literature, what the nature of the opportunities and threats are to democracy from developments in digital media.

We asked two research questions:

1. What are the specific opportunities digital media presents for improving democratic participation?
2. What are the current threats/barriers that are in place to prevent achieving those opportunities?

In total, 110 documents were reviewed (including journal articles, reports and book chapters), with 69 of those containing evidence to support one or more of the research questions (see reference list).

A non-systematic narrative review was chosen with a view to summarising the themes that have been covered in terms of opportunities and problems (risks and threats). Searches were limited to research published in the last eight years (most are within five).

SPECIFIC OPPORTUNITIES FOR IMPROVING DEMOCRATIC PARTICIPATION THROUGH DIGITAL MEDIA

This section outlines six specific opportunities provided by digital media for improving democratic participation discovered in the literature. These are: the democratisation of information publishing, the broadening of the public sphere, the increasing equality of access to and participation within political processes, increasing transparency and accountability from government and the promotion of democratic values.

2.1

DEMOCRATIZATION OF INFORMATION PUBLISHING

This is the capacity of digital media to enable “anyone to create content and share it with a global audience” (College of St George, 2018, p. 1). As well as being identified in the background paper to a consultation currently being undertaken by the UK organisation St College of St George, this capacity has been noted by the influential sociologist Manuel Castells (2013). Castells termed it “the shift of mass communication to mass self-communication” (p. 23), whereby large media corporations and governments no longer dominate the production of messages and content to the same degree as the majority of the 20th Century. Whereas in the age of mass-communication the ability to generate content was limited by access to costly printing facilities, TV studios etc., in the age of mass self-communication such entry barriers have been reduced to simply owning a laptop or a mobile phone.

This has the potential to improve democratic participation by facilitating dialogue both between governments and citizens (improving institutional trust) and between otherwise divergent groups and individuals in society.

2.2

BROADENING THE PUBLIC SPHERE

Linked to 2.1, digital media has the capacity to widen policy conversations to include “previously marginalized individuals and communities” (College of St George, 2018, p. 1), who formerly would have been excluded from democratic processes. A good example of this is the @IndigenousX Twitter account (explored in more detail in section 4.3), which provides a platform for the articulation of indigenous Australian culture and perspectives, “bringing their views and concerns to a wider audience” (Sweet, Pearson, & Dudgeon, 2013, p. 109) than would have been possible before the widespread adoption of digital media.

In a study of the US context, Auger (2013) found that social media increased the opportunities for NGOs to express their perspectives, meaning a larger “marketplace of ideas” has been able to take shape than previously. Moreover, social media meant that it was easier for non-mainstream ideas to become legitimised, which was linked to the securing of funds for the NGOs’ activities. Further, by analysing 235 NGO social media posts, Auger (2013) found that “rational appeals were the most frequent type of advocacy characteristics used” (p. 373), despite the issue studied being the highly contentious one of gun control in the US, which appears to counter the more recent emphasis on social media filter bubbles (see section 3.2). The study identified 274 different message characteristics from that corpus, assigned by the purpose, content and emphasis of posts from four different NGOs, including the National Right to Life Committee and the National Rifle Association. Only 17% of the “message appearances identified” (p. 373) fitted the study’s definition of propaganda, defined by the four features of reducing complex issues, use of authority figures, emphasis on conflict rather than cooperation and reduction of complex issues to cause and effect.

2.3

INCREASING EQUALITY OF ACCESS TO AND PARTICIPATION WITHIN POLITICAL PROCESSES

Several studies reviewed found that digital media had the potential to increase equality of access to and participation within political processes, in terms of gender, class, race and age.

In a study of the Israeli 2015 parliamentary election, Yarchi and Samuel-Azran (2018) found that Facebook afforded more positive exposure to women politicians than traditional news media. The authors found that “female politicians’ posts generated significantly more user engagement in terms of the number of Likes and Shares in comparison to male politicians” (p.978), creating a supportive communicative environment which boosted their self-esteem.

Two studies (Dubow, 2017; see Government Information Services, 2018, for the New Zealand context) found evidence through interviews with experts that digital media, when incorporated with other good government policies (such as civics education), has the potential to build well-networked, educated and empowered communities, which previously have been economically and socially marginalised by digital divides. Dubow (2017) recommended the development of new digital tools focused on breaking down and summarising civic information, while Government Information Services (2018) recommended tools which allowed different levels of participation, increasing inclusivity across genders, ethnicities and ages.

In terms of race, Jakubowicz et al. (2017) found that, through the examination of several case-studies in the Australian context, digital media can facilitate the formation of both ‘ad-hoc’ and longer-term, group-based online communities focused on fighting racism, which can provide a safe space of belonging for ethnic minority groups. This sense of community encourages engagement and participation in public discourse such as the campaign to change the date of Australia Day, which could be otherwise discouraged by online hate-speech and more mainstream forms of racist discourses (see sections 3.3 and 4.3 for further details).

A survey undertaken by the UK Think Tank Demos (Miller, 2016) found that social media and other digital media forms increased participation and engagement in the British 2015 elections by young people, an age group which is believed to have become increasingly disengaged from political processes. Similar results were derived by Xenos, Vromen, and Loader (2014), who, in a comparative study of Australia, USA and UK, found that social media can “soften traditional patterns of political inequality” (p.152), by encouraging political engagement from 16-29 year olds.

2.4 INCREASING PARTICIPATION AND ENGAGEMENT IN POLITICAL PROCESSES

Numerous studies found links between digital media and increased participation and engagement in political processes in the general population, not only in marginalised groups. This includes engagement in elections, different forms of deliberative democracy, as well as participation in more informal political action such as protests.

In terms of voting, the above mentioned Demos study (Miller, 2016) found “39 per cent of poll respondents who had engaged with political content on social media felt more likely to vote as a direct result” (p. 11). Such positive results can be partly attributed to the capacity of platforms such as Twitter to provide highly interactive, temporary political discussion fora through the hashtag function, which are easy to engage with, and accessible without high degrees of technical or political knowledge. For example, Barack Obama’s 2015 State of the Union address “spurred approximately 2.6 million tweets” (Gayo-Avello, 2015, p. 10) via the #SOTU hashtag.

In terms of participative democracy, a study of the public review of the Icelandic constitution, (Valtysson, 2013) found that the use of Facebook, Twitter, Flickr, and YouTube both increased engagement and facilitated the emergence of “networked publics” which promoted consensual opinion formation, despite lacking formal decision-making authority. In another study on the Iceland context, but in the area of local government, Simon, Bass, Boelman, and Mulgan (2017) found that the Better Reykjavik (idea generation) and Better Neighbourhoods (participatory budgeting) platforms saw 70,000 citizens taking part out of a total city population of just 120,000. The active participation of so many people would be extremely difficult to organise without digital media platforms.

The same research (a set of case studies) also looked at France’s *Parlement et Citoyens*, a “website which brings together representatives and citizens to discuss policy issues and collaboratively draft legislation” (p. 24). This initiative aims to move beyond consultation towards citizens “inform[ing] and shap[ing] legislation which is put before Parliament” (p. 24). Survey evidence from Switzerland (Kern, 2017), which has a high number of binding referendums, is that the availability of such systems of direct democracy increases feelings of having influence over the system, thereby increasing the likelihood of participation in formal democratic processes. Kern (2018), through a combination of quantitative surveys and qualitative, semi-structured interview data from Belgium, also found links between participation in a single referendum (the most common form of direct democracy) and intention to participate in future political protests.

Additionally, in the cases of the Arab Spring, Los Indignados and Occupy movements, Twitter was been found to have played a key role in the organisation of those large-scale protests (Bennett, Segerberg, & Walker, 2014; Papacharissi & de Fatima Oliveira, 2012). Particularly in the case of the Los Indignados movement in Spain (Bennett & Segerberg, 2012), digital media platforms were found to be “taking the role of established political organizations” (p. 742) such as political parties and unions, which were regarded as corrupt.

2.5 INCREASING TRANSPARENCY AND ACCOUNTABILITY FROM GOVERNMENT

A substantial amount of recent research was found on “open government” and “e-government” initiatives, with an influential OECD (2017) report highlighting their importance in rebuilding public trust in democratic institutions and policies. The general consensus is that governments which make data on the transactions of government departments available to their citizens via open government portals, allow citizens to see that taxpayer funds are being spent appropriately and fair decisions being made in a transparent manner, thereby increasing trust (see Kim & Lee, 2012; Nielsen, 2017; Wu, Ma, & Yu, 2017).

Survey research across 36 major cities in China finds that such moves towards transparency are particularly effective when overall trust in government institutions is low (Wu et al., 2017). However, a separate analysis of seven Western open-government portals finds that the “ordinary citizens” perceives terms such as “open government” and “e-government” as vague -- confusion that correlates to large differences in accessibility levels (Lourenço, 2015).

In a large quantitative study of survey data from 36 major cities in China, Wu et al. (2017) found that such moves towards transparency are particularly effective when overall trust in government institutions is low. Perceptions of equality of public service provision is also “substantially strengthened [by open government initiatives] when government trust is low (p. 898). However, Lourenço (2015) noted, through an analysis of seven Western open-government portals from the perspective of the “ordinary citizen”, that terms such as “open government” and “e-government” can be vague, allowing for large differences in levels of accessibility. Hence, if such websites do seek to enable the holding of government to account by the citizenry, they need to do more than merely dump raw data, they need to structure websites and data so that the non-data expert can use them (more on this in section 4.7).

2.6

PROMOTION OF DEMOCRATIC VALUES

The sixth and final opportunity outlined in the literature is the use and regulation of digital media by governments to actively promote democratic values, informed debate, tolerance and respect for other groups. For example, this can be done directly through government funding of public service journalism, as advocated for by the Organization for Security and Co-operation in Europe (OSCE, 2017). A second direct form of promotion can be through the funding of independent statutory organisations such as All Together Now in Australia, which is “focused generally on encouraging the embracement of cultural diversity and the eradication of racism” (Jakubowicz et al., 2017, p. 242). The organisation runs the anti-racism Twitter account @itstopswithme, together with the #itstopswithme hashtag which encourages citizen interactive engagement in campaigning.

Less direct forms are the general educative effects of participating in democratic processes, the scope and breadth of which have been shown in this section to be potentially both widened and enlarged by digital media. Michels (2011) conducted a meta-analysis on the effects of citizen participation in democratic processes, collecting “empirical evidence about effects from 120 cases in different Western countries” (p. 276). Overall findings were that government programs which promote participatory democracy have “a positive effect on the development of knowledge, skills, and virtues [which includes] active participation in public life, trustworthiness, and reciprocity (giving and taking)” (p. 278). This provides support to the OECD’s (2017, p. 118) argument that, together with short-term, functional benefits, there are also intrinsic, long-term benefits to government support for digital media-enabled participatory democracy, and will be discussed in more detail in section 4.

While the literature highlights many opportunities offered by digital media to improving democratic engagement, there are many threats also. The following section covers those revealed by the published literature.

CURRENT THREATS/BARRIERS TO ACHIEVING OPPORTUNITIES

This section will define and outline the emergent threats or barriers to increasing democratic participation through digital media. Seven key threats have been identified from the literature (there is overlap between them), which are linked to the increasing influence of digital media in society (though this is not the only cause). These threats range from issues at the structural and systems level e.g interference by governments, through to threats from individual responses, eg distrust. Together, they threaten to derail the optimistic promise of digital media through the realisation of the specific opportunities outlined in section two.

These threats have been identified as:

1. increasing power of private platforms,
2. foreign government interference in democratic processes,
3. fake news (also known as misinformation and disinformation),
4. filter bubbles (also known as echo chambers),
5. surveillance and data protection,
6. hate speech and trolling, and
7. a growing distrust of or dissatisfaction with democracy.

3.1

INCREASING POWER OF PRIVATE PLATFORMS

Private platforms have increasing power to determine all aspects of our information lives, social interactions, and democratic activities. It underpins and flows back from most of the other threats we discuss.

Sections of the reviewed scholarship highlighted the increasing dominance of an increasingly small number of privately-owned platforms over the internet (see Fuchs, 2014). Google and Facebook dominate the digital advertising market, both in the US and in New Zealand (Myllylahti, 2018; Srnicek, 2017). The two companies “drive 53 percent of [New Zealand] news websites’ traffic” (Myllylahti, 2018, p. 6), but without contributing a corresponding volume of advertising revenue, thereby threatening journalism’s economic foundations, with serious repercussions for the breadth and quality of the public sphere (opportunity 2.2).

Srnicek (2017) highlights a monopolisation tendency “built into the DNA of platforms” such as Google and Facebook, linked to the close correlation between the mining of user-data and the ability of these companies to make profits. Such is the value of data in what Srnicek terms the era of “platform capitalism”, Google and Facebook are rapidly purchasing smaller companies so that they are able to control the extraction, processing and analysis of such data, thereby setting the rules of the game, making it increasingly harder for competitors to enter the market.

Further, it becomes increasingly difficult for companies that offer alternatives to the data-extraction for profit model, such as Loomio (discussed further in Part 2), to survive in this environment (Jackson & Kuehn, 2016). Loomio is a deliberative democracy tool intentionally organised around the principles of open source (user control over source code) and co-operative, democratised ownership and decision-making. However, because it lacks the resources to sustain huge servers or cloud services required for the large amounts of data necessary for the functioning of its platform, it must lease these services from the big platforms. Hence, “it must sacrifice some aspects of control for economic reasons” (p. 424).

At the level of individual personality, our lives are led more and more through these platforms, meaning that they increasingly shape our social worlds. As put by the Internet Governance Forum (IGF, 2015), “increasingly, the operation of these platforms affects individuals’ ability to develop their own personality and engage in a substantial amount of social interactions” (p. 1). Hence, the actions of these companies can impact human rights (see also OSCE, 2017), not only through their control of personal data. Because their algorithms dictate what appears and what does not in the public sphere, their algorithms could be seen as a form of censorship. Complicating this further is that human rights protections are normally applied to national governments, rather than private companies.

3.2

FOREIGN GOVERNMENT INTERFERENCE IN DEMOCRATIC PROCESSES

A recent US intelligence report “claimed with a high degree of confidence” (Ziegler, 2018, p. 567) that “Russian President Vladimir Putin ordered an influence campaign in 2016 aimed at the US presidential election” (Intelligence Community Assessment quoted in Ziegler, 2018, p. 567). This included a deliberate strategy to use social media “to undermine confidence in the election and to magnify stories critical of Hillary Clinton” (Persily, 2017, pp. 70-71). Teams of trolls were employed in order to post negative political advertising stories online (Persily, 2017), and damaging emails were distributed through Wikileaks (Ziegler, 2018). Persily (2017) argues that the negative advertising contributed to decreased turnout and voter disengagement.

As noted by Ziegler (2018), the US has not been the only target of Russia’s military intelligence unit the GRU, with German and French elections also targeted by disinformation campaigns during 2017. Ziegler (2018) argues that such tactics should be placed in the context of a broader strategy of “hybrid warfare”, where Russia seeks to exaggerate already existing tensions and polarisations by encouraging a lack of faith in the electoral system and trust in the idea of liberal democracy.

Recently, security services in New Zealand, has revealed that New Zealand has been the target of attempts to interfere in democracy through a “range of vectors” (Moir, 2019). The threat to free and fair elections, and more general to liberal participatory and inclusive democracy through the manipulation of digital media is a well-established one.

3.3

SURVEILLANCE AND DATA PROTECTION

The 2013 exposé by Edward Snowden and the 2018 Cambridge Analytica revelations have brought the issues of data privacy and surveillance into the public eye. The former revealed that the major internet service providers were sharing the data of their customers with US government agencies such as the NSA. Further, Snowden revealed that this data collection was also being done in the other member countries of the ‘Five Eyes’ (the UK, Australia, Canada and New Zealand) in order for governments to carry out mass surveillance on their citizens (Fuchs & Trottier, 2017).

The Cambridge Analytica revelations highlighted that “Facebook gave unfettered and unauthorized access to personally identifiable information...of more than 87 million unsuspecting Facebook users to the data firm” (Isaak & Hanna, 2018, p. 56). The scandal brought to the surface the underlying mechanics of the attention economy, outlined in section 3.1. Put simply, the consumer of social media is also the product – their personal data is the oil that greases the machine, or as Ghosh and Scott (2018) put it, “behavior tracking and the business of online advertising is central to the market power of global internet platforms” (p. 6).

These revelations are having effects on the perceptions of internet users: surveys in the UK have revealed deep concerns about the such practices (Fuchs & Trottier, 2017), while a recent survey by Internet NZ revealed data security and privacy was one of the top 5 concerns of New Zealand users (InternetNZ, 2017b). In the US, surveys of writers (PEN America, 2013) and investigative journalists (Holcomb, Mitchell, & Purcell, 2015) have revealed a worrying “chilling effect”, similar to the spiral of silence mentioned earlier, which demonstrates significant curtailing effects on the private and public practices

of both professions, which are vital for the sustainment of a healthy public sphere (opportunity 2.2). Evidence from the UK indicates that ongoing surveillance of Muslim communities contributes to feelings of alienation from mainstream society (Blackwood, Hopkins, & Reicher, 2013), thereby having a detrimental effect on opportunity 2.3. In the wake of the Christchurch terror attacks in March 2019, in which 50 muslim New Zealanders were murdered by a white supremacist, surveillance of the muslim New Zealand community by government agencies through digital means is also being highlighted as a key threat to democracy (Human Rights Foundation of Aotearoa New Zealand, 2016).

3.4 FAKE NEWS/ DISINFORMATION

Probably the most famous (or infamous) of the identified threats to democratic participation, due to its links to the current US president, the term “fake news” rapidly went “from being marginal to near ubiquitous” (Farkas & Schou, 2018, p. 304) within news media discourse in the immediate lead up to and the aftermath of the November 2016 election in the US. Linked to this emergence within a highly politicised context, the term has become a tool for the delegitimization of political opponents, signalling a broader “clash of narratives” (Marda & Milan, 2018, p. 3) between conservatism and liberalism in the US.

The phenomenon could also be labelled propaganda, but as this is also a highly loaded term, disinformation is most suited to our purposes here. Disinformation is distinguished from misinformation, with the latter lacking a deliberate intention. The former, by contrast, is defined as “false or misleading information that is deliberately disseminated to deceive a target audience” (College of St George, 2018, p. 2). As well as deliberately misleading content, disinformation can be disingenuous concerning its “origins and affiliations...[often undertaking] concerted efforts to mask these origins” (FireEye, 2018, p. 5). Because it is defined by intent, disinformation can become misinformation when it is unintentionally spread by human interaction online (Vosoughi, Roy, & Aral, 2018).

Ghosh and Scott (2018) offer a yet more precise term in “political disinformation”, defined as “highly targeted political communications that reach millions of people with customized messages that are invisible to the broader public” (p. 3). Ghosh and Scott (2018) thereby link the phenomenon directly to what has been commonly termed the “attention economy”, or “the financial interests that drive the core technologies of the leading internet platforms” (p. 4). Polarising political posts (whether true or not) evoke the strongest emotions, and therefore hold attention, “which in turn generates [advertising] revenue” (p. 4). As has become clear with the recent Cambridge Analytica revelations (Isaak & Hanna, 2018), targeted political advertising is a highly profitable business.

Not only does it hold attention, but political disinformation spreads faster around the internet through likes, shares and retweets. Vosoughi et al. (2018) found that “false political news... [not only] diffused significantly farther, faster, deeper, and more broadly than the truth” (p.1), but also other highly viral types of news such as reports on terrorism. Guo, Rohde, and Wu (2018), found that this had a direct impact on the 2016 election, as “fake news sites...were mainly responsible for spreading negative news about [Hillary] Clinton” (p.14), rather than Donald Trump.

3.5

FILTER BUBBLES/ ECHO CHAMBERS

While filter bubbles and echo chambers are terms which are often used interchangeably, the former tends to refer to a specific technical effect of the attention economy, while the latter has more of a broader social and psychological dimension. The most famous example of a filter bubble is Facebook's news feed, created by a machine-learning algorithm which draws on data created by user networks, likes and comments. The algorithm (and hence the news feed) can also be influenced by and how much companies and organisations are willing to pay to be present there (hence political disinformation requires filter bubbles to be effective, see Ghosh & Scott, 2018). However, the issue is not restricted to Facebook. Political disinformation sites can also take advantage of Google's algorithm, by paying "top billing" to appear high up on searches (Ghosh & Scott, 2018). Hence, central to the profits of the two most powerful internet platforms (Srnicsek, 2017), filter bubbles are today so ubiquitous they most often work in the background to our daily lives, shaping the information we receive "imperceptibly and without consent" (College of St George, 2018, p. 5).

Filter bubbles follow a longer-term trajectory within advertising (including political advertising) which has sought to collect data in order to tailor adverts to target groups, however, now they can be targeted to specific individuals (Ghosh & Scott, 2018). This can contribute to the formation of echo chambers, which is the reinforcement of confirmation bias through selective exposure to information (College of St George, 2018; Guo et al., 2018). Hence, the technical and economic drivers of filter bubbles can act to reinforce echo chambers, but the two cannot be reduced to each other, with the latter existing before social media, through for example, the alignment of newspapers to political affiliation (Möller, Trilling, Helberger, & van Es, 2018).

Increasing numbers of automated social media 'bots' have also been linked with the spread of political disinformation and thus the reinforcement of echo chambers (Farkas & Schou, 2018; Persily, 2017). A study of Twitter during the 2016 US election between 16 September and 21 October put the number of active bots at around 400,000, which were "responsible for roughly 3.8 million tweets, about one-fifth of the entire conversation" (Bessi & Ferrara, 2016).

However, the individual user is not without agency, with the majority of false stories on Twitter still being spread, and echo chambers still being reinforced, by humans, rather than bots (Vosoughi et al., 2018). Closely linked to the reinforcement of echo chambers on Twitter is the follow function, whereby users are encouraged to follow other users who confirm similar ideological views to their own, restricting their exposure to ideologically challenging discourse (Guo et al., 2018; Himelboim, McCreery, & Smith, 2013). From a social psychology perspective, echo chambers act as an identity-securing protection from the epistemological and ontological uncertainties created by the vast amounts of (often conflicting) information available online (Lu & Yu, 2018). Linked to this is the decline in trust in the ability of traditional news media to provide reliable information (Knight Foundation, 2018).

While new research is contesting the placement of blame for echo chambers solely at the door of social media (Beam, Hutchens, & Hmielowski, 2018), there is little doubt that filter bubbles have "exacerbated political divisions and polarization" (Deb, Donohue, & Glaisyer, 2017, p. 4). This fracturing effect has negative implications for the mechanisms of liberal democracy, as developing a broad consensus around decisions made in the public good becomes increasingly difficult (OECD, 2017).

3.6

HATE SPEECH AND TROLLING

Linked to the above-mentioned fracturing effect has been the rise of hate speech and trolling. While hate speech obviously predates the internet, trolling is a term linked directly to internet cultures, and until recently, had more playful, less hurtful connotations (Phillips & Milner, 2017). Linked to the hacker breeding ground 4chan, a troll is an anonymous user who deliberately provokes antagonistic reactions for sheer enjoyment, or “the lulz” (Coleman, 2014). Trolling is aided both by the ease of creating anonymous online profiles (Galán-García, de la Puerta, Gómez, Santos, & Bringas, 2014), and by the atomised nature of internet interaction, both of which can exacerbate certain psychological profiles (Jakubowicz, 2017). A study which tallied a personality survey to one on internet use found strong correlations “between trolling and the Dark Tetrad of personality... sadism, psychopathy, and Machiavellianism” (Buckels, Trapnell, & Paulhus, 2014).

Trolling can pose a direct threat to opportunities 2.2 and 2.3 when it becomes systematically targeted towards minority groups in order to deliberately cause emotional distress, i.e. when it becomes hate speech (Alkiviadou, 2018). While not all hate speech is articulated by trolls, remaining anonymous makes individuals more likely to escape prosecution for the more egregious examples (Holschuh, 2013), as complex and time-consuming tracking systems have to be employed to trace the perpetrators (Galán-García et al., 2014). As well as

Racialised hate speech (otherwise known as cyber racism) is specifically targeted towards ethnic minority groups, and has become increasingly coordinated in recent years, through the rise of the “alt-right” (Jakubowicz, 2017). It has become a global phenomenon, affecting “refugees and ethnic minorities in Europe, Muslim Blacks and Jews in the United States, Indigenous Australians” (Jakubowicz et al., 2017, p. v) and others. It can have a direct negative impact on opportunities 2.2, 2.3 and 2.4, by encouraging affected groups to retreat to safe locations “where they focus on building intracommunal bonding” (Jakubowicz et al., 2017, p. xi), rather than engaging with national debates and institutions.

Sexualised hate speech is primarily targeted towards women (together with members of the LGBTIQI community), and is characterised by its specifically misogynistic nature (Edstrom, 2016). It is often directed towards women in the public eye, or those in influential positions, such as journalists, with proponents directing critical attention onto their supposed essential gender characteristics, rather than their work (Edstrom, 2016). Research undertaken by The Guardian newspaper on its comment threads revealed that out of the ten journalists who had attracted the most hateful comments, eight were women, with the other two being black men (Gardiner et al., 2016). This can have a negative impact on efforts towards the broadening of the public sphere, as women journalists are discouraged from writing what they may feel are controversial stories.

More generally, a large survey conducted in Hong Kong (Chen, 2018) found a correlation between strong, vocal disagreements with an individual’s perspectives and a “spiral of silence” which acts to curtail the voicing of contentious opinions by minority groups. Particularly when the polarisation of the public sphere is heightened, fear of social isolation makes it more likely that users “express less disagreeing opinion and exercise more withdrawal behaviors” (p. 3928). The particular ability of trolls and hate speech to fan antagonistic “flames” Ceron and Memoli (2016) rather than promote rational debates, therefore has a direct impact on increasing democratic participation.

3.7

DISTRUST/ DISSATISFACTION WITH DEMOCRACY

This section details a more longer-term, general process which has been ongoing since the 1960s (Ziegler, 2018), but which is intensified by the other threats already outlined in this section. This is increasing distrust and/or dissatisfaction with democratic processes, together with government institutions and politicians. Emblematic of this decline in confidence is a 2016 Pew Research survey which found that trust in the US national government had hit an historic low at 20 percent (Ziegler, 2018). While New Zealanders' trust in their public services was a little higher at 42 percent the last time the data was collected in 2015 (Stats NZ, 2015), a general, widespread decline in levels of trust in government is acknowledged to be a major issue affecting a majority of the wealthier, Western nations (OECD, 2017).

Deb et al. (2017) contend that “that fundamental principles underlying democracy—trust, informed dialogue, a shared sense of reality, mutual consent, and participation—are being put to the test by certain features and attributes of social media” (p. 3). These include the aforementioned “echo chambers”, which, when combined with the “proliferation of partisan media in traditional channels, has exacerbated political divisions and polarization” (p. 4).

Lu and Yu (2018) found, drawing on the World Values Survey 2010-14, that the “decentralized, horizontal, interactive mode of information distribution” (p.3) characterised by the internet, increases intolerance of others (although they also found that participation in public deliberation with others from outside their echo chambers increases tolerance – more on this in section 4). More specifically, Ceron and Memoli (2016), drawing on a Eurobarometer survey which collects data from 27 European countries, found that it is the consumption of online forms of news that include opinions which differ from the attitudes of the consumer, which increase “the likelihood of ‘flames’ [a strong polarising effect] that increase skepticism toward democracy” (p.226).

Hence these findings appear to go against the assumption, embedded within the “optimistic” perspective (see Castells, 2013), that increasing access to varied information and differing opinions automatically facilitates a more open, tolerant and inclusive society.

CONCLUSION

In this narrative literature review we outlined a cross-section of the recent international literature on the key opportunities for the expansion of digital democracy, and the current threats to actualising those opportunities.

The six key opportunities identified were: the democratisation of information publishing, the broadening of the public sphere, the increasing equality of access to and participation within political processes, increasing transparency and accountability from government and the promotion of democratic values.

These opportunities are at risk however from a significant group of threats, all of which are interconnected to the major structural threat posed by the increasing power of private platforms. The other six threats identified are: foreign government interference in democratic processes, surveillance and data protection, fake news (also known as misinformation and disinformation), filter bubbles (also known as echo chambers), hate speech and trolling, and a growing distrust of or dissatisfaction with democracy.

To ensure that the potential of digital media technologies is realised in relation to digital democracy, it is important to understand what the research says works to limit or overcome these threats. This is the focus of part two of this literature review.

REFERENCES

- Alkiviadou, N. (2018). Hate speech on social media networks: towards a regulatory framework? *Information & Communications Technology Law*, 1-17. doi:10.1080/13600834.2018.1494417
- Auger, G. A. (2013). Fostering democracy through social media: Evaluating diametrically opposed nonprofit advocacy organizations' use of Facebook, Twitter, and YouTube. *Public Relations Review*, 39(4), 369-376. doi:https://doi.org/10.1016/j.pubrev.2013.07.013
- Beam, M. A., Hutchens, M. J., & Hmielowski, J. D. (2018). Facebook news and (de) polarization: reinforcing spirals in the 2016 US election. *Information, Communication & Society*, 21(7), 940-958. doi:10.1080/1369118X.2018.1444783
- Bennett, W. L., & Segerberg, A. (2012). The logic of connective action: Digital media and the personalization of contentious politics. *Information, Communication & Society*, 15(5), 739-768. doi:10.1080/1369118X.2012.670661
- Bennett, W. L., Segerberg, A., & Walker, S. (2014). Organization in the crowd: peer production in large-scale networked protests. *Information, Communication & Society*, 17(2), 232-260. doi:10.1080/1369118X.2013.870379
- Bessi, A., & Ferrara, E. (2016). Social bots distort the 2016 US Presidential election online discussion. *First Monday*, 21(11). doi:https://doi.org/10.5210/fm.v21i11.7090
- Blackwood, L., Hopkins, N., & Reicher, S. (2013). Turning the Analytic Gaze on "Us": The Role of Authorities in the Alienation of Minorities. *European Psychologist*, 18(4), 245-252. doi:10.1027/1016-9040/a000151
- Buckels, E. E., Trapnell, P. D., & Paulhus, D. L. (2014). Trolls just want to have fun. *Personality and Individual Differences*, 67, 97-102. doi:https://doi.org/10.1016/j.paid.2014.01.016
- Castells, M. (2013). *Communication power (Second ed.)*. Oxford: Oxford University Press.
- Ceron, A., & Memoli, V. (2016). Flames and Debates: Do Social Media Affect Satisfaction with Democracy? *Social Indicators Research*, 126(1), 225-240. doi:10.1007/s11205-015-0893-x
- Chen, H.-T. (2018). Spiral of silence on social media and the moderating role of disagreement and publicness in the network: Analyzing expressive and withdrawal behaviors. *New media & society*, 20(10), 3917-3936. doi:10.1177/1461444818763384
- Coleman, G. (2014). *Hacker, Hoaxer, Whistleblower, Spy: The Many Faces of Anonymous*. London: Verso.

- College of St George. (2018). *Democracy in a Post-Truth Information Age: Background Paper*. Retrieved from Windsor: <https://www.stgeorghouse.org/wp-content/uploads/2017/10/Background-Paper.pdf>
- Deb, A., Donohue, S., & Glaisyer, T. (2017). Is Social Media a Threat to Democracy? *Omidyar Group*. <https://www.omidyargroup.com/wp-content/uploads/2017/10/Social-Media-and-Democracy-October-5-2017.pdf>.
- Dubow, T. (2017). *Civic Engagement: How Can Digital Technologies Underpin Citizen-Powered Democracy?* Retrieved from https://www.rand.org/pubs/conf_proceedings/CF373.html
- Edstrom, M. (2016). The trolls disappear in the light: Swedish experiences of mediated sexualised hate speech in the aftermath of Behring Breivik. *International Journal for Crime, Justice and Social Democracy*, 5(2), 96-106.
- Farkas, J., & Schou, J. (2018). Fake News as a Floating Signifier: Hegemony, Antagonism and the Politics of Falsehood. *Javnost - The Public*, 25(3), 298-314. doi:10.1080/13183222.2018.1463047
- FireEye. (2018). *Suspected Iranian influence operation: Leveraging inauthentic news sites and social media aimed at U.S., U.K., other audiences*. Retrieved from <https://www.fireeye.com/content/dam/fireeye-www/current-threats/pdfs/rpt-FireEye-Iranian-IO.pdf>
- Fuchs, C. (2014). Social Media and the Public Sphere. *tripleC*, 12(1), 57-101.
- Fuchs, C., & Trottier, D. (2017). Internet surveillance after Snowden: A critical empirical study of computer experts' attitudes on commercial and state surveillance of the Internet and social media post-Edward Snowden. *Journal of Information, Communication and Ethics in Society*, 15(4), 412-444. doi:10.1108/JICES-01-2016-0004
- Galán-García, P., de la Puerta, J. G., Gómez, C. L., Santos, I., & Bringas, P. G. (2014, 2014//). *Supervised Machine Learning for the Detection of Troll Profiles in Twitter Social Network: Application to a Real Case of Cyberbullying*. Paper presented at the International Joint Conference SOCO'13-CISIS'13-ICEUTE'13, Cham.
- Gardiner, B., Mansfield, M., Anderson, I., Holder, J., Louter, D., & Ulmanu, M. (2016). The dark side of Guardian comments. *The Guardian*. Retrieved from <https://www.theguardian.com/technology/2016/apr/12/the-dark-side-of-guardian-comments>
- Gayo-Avello, D. (2015). Social Media, Democracy, and Democratization. *IEEE MultiMedia*, 22(2), 10-16. doi:10.1109/MMUL.2015.47
- Ghosh, D., & Scott, B. (2018). *#digitaldeceit: The Technologies Behind Precision Propaganda on the Internet*. Retrieved from <https://www.newamerica.org/public-interest-technology/policy-papers/digitaldeceit/>
- Government Information Services. (2018). *How digital can support participation in government*. Retrieved from Wellington: <https://www.digital.govt.nz/standards-and-guidance/engagement/online-engagement/research-how-digital-can-support-participation-in-government/>

- Guo, L., Rohde, J. A., & Wu, H. D. (2018). Who is responsible for Twitter's echo chamber problem? Evidence from 2016 U.S. election networks. *Information, Communication & Society*, 1-18. doi:10.1080/1369118X.2018.1499793
- Himelboim, I., McCreery, S., & Smith, M. (2013). Birds of a Feather Tweet Together: Integrating Network and Content Analyses to Examine Cross-Ideology Exposure on Twitter. *Journal of Computer-Mediated Communication*, 18(2), 40-60. doi:10.1111/jcc4.12001
- Holcomb, J., Mitchell, A., & Purcell, K. (2015). Investigative journalists and digital security. *Pew Research Center*. Retrieved from <http://www.journalism.org/2015/02/05/investigative-journalists-and-digital-security>
- Holschuh, J. (2013). #CivilRightsCybertorts: Utilizing Torts to Combat Hate Speech in Online Social Media. *University of Cincinnati Law Review*, 82, 953.
- Human Rights Foundation of Aotearoa New Zealand. (2016). And you think maybe this is not home. *Human Rights Foundation of Aotearoa New Zealand*. Retrieved from <https://humanrightsfoundation.files.wordpress.com/2019/03/and-you-think-maybe-this-is-not-home-final.pdf>
- IGF. (2015). *Recommendations on terms of service and human rights*. Retrieved from <https://www.intgovforum.org/cms/documents/igf-meeting/igf-2016/830-dcpr-2015-output-document-1/file>
- InternetNZ. (2017b). *What kiwis think of the Internet in New Zealand*. Retrieved from <https://internetnz.nz/2017-internet-research>
- Isaak, J., & Hanna, M. J. (2018). User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection. *Computer* (8), 56. doi:10.1109/MC.2018.3191268
- Jackson, S. K., & Kuehn, K. M. (2016). Open Source, Social Activism and "Necessary Trade-offs" in the Digital Enclosure: A Case Study of Platform Co-operative, Loomio.org. *tripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society*, 14(2), 413-427.
- Jakubowicz, A. (2017). Alt_Right White Lite: trolling, hate speech and cyber racism on social media. *Cosmopolitan Civil Societies: An Interdisciplinary Journal*, 9(3), 41.
- Jakubowicz, A., Dunn, K., Mason, G., Paradies, Y., Bliuc, A.-M., Bahfen, N., Oboler, A., Atie, R., Connelly, K. (2017). *Cyber Racism and Community Resilience: Strategies for Combating Online Race Hate*: Springer, 5-11, 242. doi:10.1007/978-3-319-64388-5
- Kern, A. (2017). The effect of direct democratic participation on citizens' political attitudes in Switzerland: The difference between availability and use. *Politics and Governance*, 5(2), 16-26. doi:<http://dx.doi.org/10.17645/pag.v5i2.820>
- Kern, A. (2018). What happens after a local referendum? The effect of direct democratic decision-making on protest intentions. *Local Government Studies*, 44(2), 183-203. doi:10.1080/03003930.2017.1411809
- Kim, S., & Lee, J. (2012). E-Participation, Transparency, and Trust in Local Government. *Public Administration Review*, 72(6), 819-828. doi:10.1111/j.1540-6210.2012.02593.x

- Nielsen, M. M. (2017). eGovernance frameworks for successful citizen use of online services: A Danish-Japanese comparative analysis. *JeDEM-eJournal of eDemocracy and Open Government*, 9(2), 68-109.
- Knight Foundation. (2018). *American views: trust, media and democracy*. Retrieved from <https://knightfoundation.org/reports/american-views-trust-media-and-democracy>
- Leininger, A. (2015). Direct Democracy in Europe: Potentials and Pitfalls. *Global Policy*, 6(S1), 17-27. doi:10.1111/1758-5899.12224
- Lourenço, R. P. (2015). An analysis of open government portals: A perspective of transparency for accountability. *Government Information Quarterly*, 32(3), 323-332. doi:<https://doi.org/10.1016/j.giq.2015.05.006>
- Lu, J., & Yu, X. (2018). Does The internet make us more intolerant? A contextual analysis in 33 countries. *Information, Communication & Society*, 1-15. doi:10.1080/1369118X.2018.1499794
- Marda, V., & Milan, S. (2018). Wisdom of the Crowd: Multistakeholder Perspectives on the Fake News Debate. *Internet Policy Review Series, Annenberg School of Communication*, 3.
- Michels, A. (2011). Innovations in democratic governance: how does citizen participation contribute to a better democracy? *International Review of Administrative Sciences*, 77(2), 275-293. doi:10.1177/0020852311399851
- Miller, C. (2016). *The rise of digital politics*. Retrieved from <https://www.demos.co.uk/project/the-rise-of-digital-politics/>
- Moir, J. (2019). Foreign political interference in NZ from ‘a range of actors’ - spy agency heads. *Radio New Zealand*. Retrieved from <https://www.radionz.co.nz/news/political/386865/foreign-political-interference-in-nz-from-a-range-of-actors-spy-agency-heads>
- Möller, J., Trilling, D., Helberger, N., & van Es, B. (2018). Do not blame it on the algorithm: an empirical assessment of multiple recommender systems and their impact on content diversity. *Information, Communication & Society*, 21(7), 959-977. doi:10.1080/1369118X.2018.1444076
- Myllylahti, M. (2018). *Google, Facebook and New Zealand news media*. Retrieved from Auckland, New Zealand: <https://thepolicyobservatory.aut.ac.nz/publications/google-facebook-and-new-zealand-news-media>
- OECD. (2017). *Trust and public policy: How better governance can help rebuild public trust*. Paris: OECD Publishing.
- OSCE. (2017). *Joint declaration on freedom of expression and “fake news”, disinformation and propaganda*. Retrieved from <https://www.osce.org/fom/302796>
- Papacharissi, Z., & de Fatima Oliveira, M. (2012). Affective News and Networked Publics: The Rhythms of News Storytelling on #Egypt. *Journal of Communication*, 62(2), 266-282. doi:10.1111/j.1460-2466.2012.01630.x
- PEN America. (2013). *Chilling effects: NSA surveillance drives US writers to self-censor*. New York: PEN American Center.

- Persily, N. (2017). The 2016 US Election: Can democracy survive the internet? *Journal of democracy*, 28(2), 63-76.
- Phillips, W., & Milner, R. M. (2017). *The ambivalent internet: Mischief, oddity, and antagonism online*. Cambridge, UK: Polity.
- Simon, J., Bass, T., Boelman, V., & Mulgan, G. (2017). *Digital democracy: the tools transforming political engagement*. Retrieved from <https://www.nesta.org.uk/report/digital-democracy-the-tools-transforming-political-engagement/>
- Srnicek, N. (2017). *Platform capitalism*: Cambridge, UK: Polity Press.
- Stats NZ. (2015). Trust in government institutions. *New Zealand Government*. Retrieved from http://archive.stats.govt.nz/browse_for_stats/snapshots-of-nz/nz-social-indicators/Home/Trust%20and%20participation%20in%20government/trust-govt-instit.aspx
- Sweet, M., Pearson, L., & Dudgeon, P. (2013). @Indigenousx: A Case Study of Community-Led Innovation in Digital Media. *Media International Australia*, 149(1), 104-111. doi:10.1177/1329878x1314900112
- Valtysson, B. (2013). Democracy in disguise: the use of social media in reviewing the Icelandic Constitution. *Media, Culture & Society*, 36(1), 52-68. doi:10.1177/0163443713507814
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146-1151.
- Wu, W., Ma, L., & Yu, W. (2017). Government Transparency and Perceived Social Equity: Assessing the Moderating Effect of Citizen Trust in China. *Administration & Society*, 49(6), 882-906. doi:10.1177/0095399716685799
- Xenos, M., Vromen, A., & Loader, B. D. (2014). The great equalizer? Patterns of social media use and youth political engagement in three advanced democracies. *Information, Communication & Society*, 17(2), 151-167. doi:10.1080/1369118X.2013.871318
- Yarchi, M., & Samuel-Azran, T. (2018). Women politicians are more engaging: male versus female politicians' ability to generate users' engagement on social media during an election campaign. *Information, Communication & Society*, 21(7), 978-995. doi:10.1080/1369118X.2018.1439985
- Ziegler, C. E. (2018). International dimensions of electoral processes: Russia, the USA, and the 2016 elections. *International Politics*, 55(5), 557-574. doi:10.1057/s41311-017-0113-1

APPENDIX 2: LITERATURE REVIEW PART 2

D I G I T A L T H R E A T S T O D E M O C R A C Y



> > L I T E R A T U R E >

> > R E V I E W > > > >

> > > > P A R T > 2 : > >

> S O L U T I O N S > > >



> > Kathleen Kuehn, Leon Salter, Jess Berentson-Shaw & Marianne Elliott



Luminate

D I G I T A L T H R E A T S T O D E M O C R A C Y

This report is part of the Digital Threats to
Democracy research project.

To see the rest of the reports and the
overall findings go to digitaldemocracy.nz

ISBN: 978-0-473-48026-4

CONTENTS

PART 1. WORKABLE SOLUTIONS	97
Decreasing the Power of Private Platforms	98
Foreign government interference in democratic processes	99
Surveillance & Data Protection	99
Fake News	100
Filter Bubbles/Echo Chambers	100
Hate Speech/Trolling	102
Distrust/dissatisfaction with democracy	106
PART 2. A SUMMARY OF COMMON RECOMMENDATIONS ACROSS THE LITERATURE	108
PART 3. INTERVENTIONS TRIED IN NEW ZEALAND	111
CONCLUSIONS	113
REFERENCES	115

Following on from the review of the literature identifying the opportunities and threats that digital media pose to an inclusive and participatory democracy (Part One), we undertook a review to identify tested and workable solutions to realising the potential of digital media and/or overcoming current threats. Underlying this work is an understanding that ensuring an inclusive and fully participatory democracy is of such critical importance to our society that the current threats posed to our democracy from digital media requires us to understand what strategies, policies and behaviours we can use to achieve such an outcome and to act.

The review is presented in three parts: 1) the empirical evidence on workable solutions to threats to democracy from digital media, 2) a summary of recommendations found in the literature and 3) a brief discussion of some activities identified in New Zealand

A non-systematic narrative review was chosen with a view to summarising the evidence. Searches were limited to research published in the last eight years (most are within five). It was not an exhaustive review, but in general we found a dearth of empirically tested solutions. This dearth of tested interventions is not surprising given the slow response of government and other public institutions (from where such research would most logically be situated and or funded) to the threats from digital media.

PART 1.

WORKABLE SOLUTIONS

In this part of the paper we present the interventions and solutions we identified in relation to each of the problems and threats that were identified in Part One of this literature review. We start with decreasing the power of private platforms.

DECREASING THE POWER OF PRIVATE PLATFORMS

Unfortunately little empirical evidence or evidence-based research has established what works for decreasing the power of private intermediaries. The dominant recommendations are generally calls for 1) more regulation (Betkier, 2018; Gillespie, 2018; Kamara, 2017; Marda & Milan, 2017); and 2) citizen-consumer activism, which ranges from pro-democracy to anti-capitalist initiatives (Srnicek, 2017; Zuboff, 2019).

The process for enacting these workable solutions also remain, at present, vague and non-specific. The suggestions are typically based on conclusions reached from multi-stakeholder interviews and focus groups as prescribed by legal scholars, academics, activists and industry leaders working in this space. Examples of recommendations to combat platform capitalism include:

Regulate platforms like other industries. Currently, regulatory debates largely centre around defining the structure, terms and conditions of what kind of industry private intermediaries represent. How platforms should be regulated or governed thus partly hinges on how these services are defined; for example, whether social media platforms are media companies, public spaces, utilities or some other service largely informs how they can ultimately be governed.

Introduce new modes of collective action (Zuboff, 2019). Under industrial capitalism we had collective bargaining, the strike – e.g., forms of collective action that were sanctioned by law and had support of society that allowed people to tame capitalism with legal protection. New forms of collective, collaborative action that connect users/consumers with the market and state to tame and outlaw surveillance capitalism must be invented. Zuboff's recommendations here map onto the work being done by advocacy groups like OHPI and autonomist Marxists (Hardt & Negri; Lazzarato; Virilio), who have been working to engage multiple vested interests lobbying for broader structural changes across the political economy *and* culture.

Creating a “sea change in public opinion” (Zuboff, 2019) that will no longer tolerate “surveillance capitalism” as the dominant economic form that trades in human futures, or one where government dips into servers held by private intermediaries for surveillance purposes. Zuboff believes a sleeping democracy has allowed these companies to create asymmetries of knowledge and power antithetical to democracy and that a shift in public attitudes is needed to persuade these companies to change.

Design New Competitive Solutions (Jackson & Kuehn, 2016; Zuboff, 2019). Dissatisfaction with large intermediaries' power presents new business opportunities that can forge an alternative digital future that in turn, facilitates a more democratic internet. Zuboff believes growing public dissatisfaction provides a space for disruptive technology to emerge; that is, platforms offering a different set of techno-social affordances have an untapped market waiting to be exploited. Platform cooperatives like Loomio (Jackson & Kuehn 2016; Ombler et al., 2016; Sandoval, 2016; Scholz, 2016; Stohl et al., 2018), subscription-based models and pro-privacy and non-commercial alternatives (Beattie, forthcoming) are already in use but have not reached the critical mass needed to become a competitive solution just yet.

Improve Content Moderation (see ‘Recommendations’ section below).

FOREIGN GOVERNMENT INTERFERENCE IN DEMOCRATIC PROCESSES

The research in this area is also largely normative, but seems to generally prescribe the design of new anti-cybersecurity infrastructure that will reduce threats to elections and other political processes (Hicks, 2018; Shoorbajee, 2018). Drawing upon “big datasets” to review and assess electoral policies, processes and outcomes – including the potential interference from foreign governments – is another common recommendation (The Election Administration and Voting Survey). Generally, most call for multi-stakeholder input on solution design, but proposals largely seem to emphasize technical over human solutions.

SURVEILLANCE & DATA PROTECTION

The opacity of mass and commercial surveillance presents challenges to empirical research looking to find and test workable solutions to the threat it poses to democracy. Some regulatory measures, like the Singaporean Data Protection Act 2012, work to regulate companies’ information management practices and have been proven effective in bringing formal charges to data mismanagement and abuse. Singapore’s regulatory measure created a Data Protection Authority (DPA) that it invested with the power to: conduct investigations on the data collection policies and practices of organisations; order the destruction of data; and impose fines for data mismanagement and abuse (Lanois, 2016). (The act extends to companies that are not based in Singapore, as well). By 2016, the DPA had received 667 complaints, took action against 11 organizations for data breaches and lodged four financial penalties (Lanois, 2016). One particular prosecution against smartphone manufacturer Xiaomi resulted in the company making changes to the terms of its cloud-messaging service agreement and data storage practices.

A majority of research, however, calls for regulatory changes to data privacy policies (e.g., Fuchs & Trottier, 2017; Internet Governance Forum, 2015; Flew, 2019). However, there is little evidence to suggest that these changes will reduce surveillance/data collection so much as regulate how that data is stored, accessed and used by data collectors and other third parties. Instead, privacy advocates typically encourage internet users concerned with reducing online surveillance or mitigating the anxieties and concerns it engenders by employing technical solutions like ad-blockers and ad-tracking browser extensions, private browser options (e.g. Tor), open source platforms and cooperative platform models all present alternatives to subvert surveillance mechanisms (Narayanan & Reisman, 2017). “Evidence” supporting the efficacy of these tools and alternatives, however, is typically anecdotal or prescriptive in nature (as opposed to empirical), and is based on the premise that these alternatives increase public awareness of tracking, bring it into public debate and in turn, decrease some of the more invasive surveillance practices.

FAKE NEWS

Few studies on fake news have yet to offer concrete empirical solutions for combatting this phenomenon. The dominant mode of published research is largely normative in approach, making prescribed recommendations for potential solutions to combatting fake news, yet none of which have been tried or tested. Methodologically, these recommendations are typically the result of multi-stakeholder interviews with experts across the public and private sectors (e.g., Marda & Milan, 2018) or joint declarations from international organisations (e.g., OSCE, 2017). Examples of proposed – but untested – recommendations to combat fake news starting at the structural level and moving towards individual level include:

- Supporting a vibrant and diverse media sphere that balances strong, independent and adequately resourced public service media with a non-concentrated commercial media sector (OSCE, 2017);
 - Developing and circulating persuasive counter-narratives [with emotional, versus rational, appeal] (Dubow, 2017; OSCE, 2017)
 - A multi-stakeholder approach to content moderation that combines human and technical intervention (Klonick, 2018; Madra & Milan, 2018)
 - Education, particularly around critical thinking (Dubow, 2017)
-

FILTER BUBBLES/ ECHO CHAMBERS

PLATFORM DESIGN & AFFORDANCES

Technically speaking, affordances refer to the perceived range of possible actions related to the features of any given platform (Bucher & Helmond, 2018; Hutchby, 2001). In nontechnical language design affordances are invitations, prompts and clues on how to use a thing. In digital media they ascribe meaning to how to use the tool, for example Facebook has a “friend” button directing the user towards ways of interacting based on mutual agreement, also a “share” button, while Twitter has a “follow” button, open to all people using the platform, directing or suggesting different ways of interacting. Other examples includes the use of anonymous accounts, retweet buttons, mentions or share data feedback. The design of these affordances has an impact on inclusion and participation, as well as the types of interactions people experience and information they are exposed to.

There is some suggestion that design affordances can reduce the effects of filter bubbles by engaging internet users in more ideologically diverse communities. While designing features to encourage or delimit certain behaviours or activities in an online space does not guarantee they’ll be taken up that way by users (and in fact, users often take up features in ways unintended or unanticipated by designers and platform owners), platform features have also been shown to effectively direct or delimit action in certain ways (Bucher & Helmond, 2018).

Non-commercial platforms like Loomio, for example, afford different modes of interaction based on the features (e.g., tools, interface) and environment (e.g., deliberative; asynchronous) it makes available outside a commercial space. Unlike privately owned social media platforms, user identities are less curated towards a consumerist framework,

rather the platform affords an environment of less performative, and thus more considered, dialogue, discussion and debate. The platform's design at least partly contributes to its effective use by government agencies, local councils and advocacy groups to consult and deliberate with citizens on a range of contentious initiatives and strategies.

The relationship between design and civility is demonstrated in a study that analysed The Wellington City Council's use of Loomio to debate proposed changes to alcohol licensing rules, which the Council ruled as a success in more creating in-depth discussion with less polarization than offline debates. Getting constituents to meaningfully engage outside echo chambers may be partly tied to Loomio's time-based affordances. The limited time given to public face-to-face meetings often creates pressure for speakers to get their points across forcefully and enhance selective bias by creating discursive alliances between constituents who seemingly share similar opinions. Loomio, however, "opens up space where all views can be considered and everyone can still be heard" without time-based anxieties and restrictions (Ombler, Russell, & Rivera-Munoz, 2016, p. 23). Loomio's platform affordances also move deliberation beyond *debate* to collective *agreement* (i.e., rational consensus) with the goal of making "the fewest people unhappy" (Rushkoff, 2014, n.p.).

These kind of well-designed, collectively-owned, online deliberative fora not only offer a safe place for different views to be expressed and heard (effectively reducing the propensity to engage with similar-minded people encouraged by automated filter bubbles), but offers a safe space for marginalized groups, as well. Loomio was effectively used by Stats NZ to debate the inclusion of "gender diverse" as a third category.¹ Despite criticisms surrounding the process, an analysis of feedback left on the Stats NZ website reported the process as largely positive, which contributors "recognized to be an inclusive discussion" (Stohl, Stohl & Ganesh, 2018, p. 246). Further, it gave "visibility to an issue that needed to be made more public" and "brought people together into the conversation who had previously been marginalized" (p. 246). As detailed in Part 1, hate speech instigated by members of a vocal majority can curtail the voicing of opinions by minority groups, pushing them further into the margins, thereby narrowing the public sphere.

1. The decision which was later rescinded, however, "for purely statistical reasons" (Stats NZ, 2018).

HATE SPEECH/ TROLLING

IDENTITY VERIFICATION SYSTEMS

While hate speech and trolling actually constitute a small minority of online comments (Papacharissi, 2004; Rowe, 2015), attempts to combat intolerance, incivility and other forms of disruptive speech have a long history in internet research. To date, however, there little consensus on efficacious strategies.

The role of anonymity online has been perhaps one of the most commonly sites of analysis for mitigating hateful, uncivil and impolite behaviour online. While early scholars advocated on behalf of anonymity's liberating affordances, namely around identity play and self-expression (e.g., Baym, 2012), the dominant view in empirical research positions anonymity as directly related to increased incivility online. Trolling, for example, generally hinges on abuse by anonymous individuals with "fake" accounts.

Sites that force pre-registration have been shown to solicit qualitatively better, but quantitatively fewer, user comments because of the extra effort required for engaging in discussion; forced registration can also facilitate 'known' user identities, which can serve to hold users to account (Bakker, 2010; Rowe, 2015; Santana, 2014). Quantitatively, when comparing user comments posted to the same stories on a news outlet's website and its corresponding Facebook page, Internet users are far more likely to comment on the website version of news stories than similar stories featured on the organisation's Facebook page (in both studies, website comments enabled anonymous posting, whereas Facebook does not).

Empirical research has also found that abusive comments are minimised when anonymous commenting is prohibited (Santana, 2014) or when identity verification is required through systems like Facebook comments (Rowe, 2015). One study, however, found that despite the ability to remain anonymous, comments on news websites were of comparatively higher quality to those on Facebook and sparked more vibrant debate, whereas Facebook comments contributed little engagement; the authors concluded that "Facebook will provide few comments, will kill the trolls, but will not result in making the conversation more interesting" (Hille & Bakker, 2014, p. 572).

Experimental research similarly shows that identification-based systems like Facebook comments correlate to more civil forms of online discourse than on platforms in which comments are posted anonymously on the same news stories (Rowe, 2015). Rowe compared news comments posted to the same stories featured on *Washington Post's* website to comments posted to the newspaper's Facebook page and found the rate of uncivil behaviour to be significantly higher on the former (where comments were anonymous) than on the its Facebook page (where comments are not anonymous). The study also found incivility to be overall more personally insulting and directed at specific users participating in political discussion on the newspaper's website. On Facebook, however, incivility and impoliteness were "aimed at individuals not involved in the discussion, or used as a way to articulate an argument, rather than offend others" (p.132). Rowe concluded that visibility to one's wider social network afforded by Facebook ostensibly holds users to account for their communicative actions in ways that website's anonymity does not.

Facebook's affordance of social surveillance thus functions as a sanctioning mechanism that enables more civil and tolerant political discussions. The conclusion, then, is that anonymity encourages more abusive behaviour while visibility or forced identification can enable more civilised forms of discourse.

PLATFORM DESIGN & AFFORDANCES

De-anonymising internet communications speaks to the role that platform design can play in mitigating incivility and hate speech online. A system or platform's communicative and social affordances at least partly explain how and why some cultures of participation evolve the way they do across different social media platforms. While user characteristics, user goals/objectives and cultural context also significantly influence how platforms are engaged, there is some evidence that suggests well-designed, digital fora can encourage stronger deliberation practices, mutual respect and accountability through their discursive promotion of implicit and explicit rules of conduct (Kavada, 2012). Research suggests that intentionally building more participatory forms of engagement into platforms might reduce filter bubbles, echo chambers *and* incivility (particularly on mobile devices), while increasing communication and deliberative processes (Groshek & Cutino, 2016; Halpern & Gibbs, 2012; Hmielowski, Hutchens, & Cicchirillo, 2014).

Therefore, the act of *consciously* designing social platforms to engender pro-social forms of engagement can have a demonstrated impact on civility. However, so also is the act of *consciously selecting and choosing* a particular platform or set of tools over others; for e-government initiatives in particular (see section on distrust/dissatisfaction with democracy), carefully selecting what (and how) social media platforms are engaged in relation to the overall objective can assist organisations, governments and other actors in facilitating more deliberative, civil forms of online discourse, as well. Writing code (e.g., features; interface design; verification systems) that enables pro-social forms of engagement and intentionally disables or limits anti-social behaviours and actions is thus one workable solution that engineers can take up as they develop and innovate new communicative platforms.

CIVIL SOCIETY & ADVOCACY ORGANISATIONS: BUILDING RESILIENCE THROUGH SUPPORT NETWORKS

Developing fast and effective reporting mechanisms and support networks are another set of workable solutions thought to combat hate speech online. Advocacy and civil society organisations like All Together Now, the Federation of Ethnic Communities' Councils of Australia (FECCA), and Australia's Online Hate Prevention Institute (OHPI) have demonstrated some success with building online reporting tools that rely on crowdsourcing to identify – in order to remove - racist hate speech online (Bodkin-Andrews, Newey, O'Rourke, and Craven, 2013, Jakubowicz et al., 2017, Sweet et al., 2013; Oboler & Connolly, 2018). For example, the OHPI's website FightAgainstHate.com offers a reporting tool that gathers and compiles evidence on the extent of the issue while also highlighting moderation gaps amongst larger online platforms.

Elsewhere, case studies of online communities like @IndigenousX (a "grassroots Twitter community made up of Indigenous Australian guest tweeters and followers") have also shown that a networked approach can effectively combat the effects of hate speech; in this case, @IndigenousX exemplifies a growing base of "resilient communities" that act as support networks to individuals who come under racist attack (Jakubowicz et al., 2017, p. 236). These groups can effectively use Twitter as a form of "participatory journalism" to build counter-narratives that counteract racism directed at indigenous Australians or others (Sweet et al, 2013).

Other research has shown the role that Twitter hashtags play in providing a space for the construction of resistant collective identities in cases like the #BlackLivesMatter movement or #MeToo (Ray, et al., 2017). The extent to which victims of online abuse report higher levels of ‘resilience to racism’ through these tactics and support networks, however, has not been empirically tested to our knowledge.

Coordinating diverse stakeholders to apply pressure to private intermediaries, particularly in ‘long-haul’ campaigns, has also been effective in having hateful content removed from social media, even when content does not technically violate a platform’s terms of service (ToS). Facebook eventually conceded to the removal of content and imagery expressing hatred against Aboriginal Australians after pressure from the OPHI, not because the content violated Facebook’s ToS but because a broad and diverse range of actors found the content insulting and complained, which included civil society, advocacy groups, regulators and individual users. Sustained pressure from diverse stakeholders tends to garner mainstream media attention that threatens the brand reputation of platforms can have a demonstrated effect on the forced removal and moderation of hateful online content (“Aboriginal Memes & Online Hate”, 2012; Gargliardone et al., 2015). Speed of removal is considered essential to diffusing the power of hate speech and trolling, as the longer hateful content remains online, the more damage it inflicts on victims while empowering the perpetrators (OPHI). Early content removal thus effectively limits the scale of exposure.

Pressure from researchers and advocacy groups alike have also encouraged some platforms to design more pro-social tools (i.e., affordances) into their systems. Twitter has responded to public pressure by unrolling a number of features that afford users the opportunity and ease to contribute to content moderation. These features include the ability to hide content users do not wish to see; the ability to flag harmful or inappropriate content (“Report Abuse” buttons); verified accounts aimed at building trustworthiness; a “quality” filter that uses an algorithm to hide spam-like tweets; a “notifications” filter for disabling notifications of mentions or replies from people they don’t follow and so on (Klonick, 2016). While these features afford users the opportunity to intentionally or semi-automatically moderate harmful and uncivil content, they also threaten to oversanitise online spaces, intensify filter bubbles/echo chambers and to disconnect users from wider network affordances (e.g., diversity, new followers, etc).

CONTENT MODERATION POLICIES & PRACTICES

Calls for new regulatory policies around content moderation at large intermediaries are gaining traction as a necessary means of combatting both incivility and misinformation online. This remains the case even in light of the growing consensus that content moderation remains an opaque and difficult practice, and on its own is not a fix-all solution.

On the one hand, the internet’s global reach presents the basic challenge of moderating what constitutes “appropriate” content across vastly different cultures marked by their own distinct values and standards, which create variance across how even ‘objective’ content rules and policies are interpreted and instituted (Roberts, 2016). Contextual issues also make quantifying inappropriate content difficult; e.g., resolving the debate over why one piece of content is acceptable but a slight variation breaks policy in another is the “holy grail” of moderation (Diakopoulos & Naaman; Klonick, 2018; Pöyhtäri, 2014).

On the other hand, content moderation policies at the largest intermediaries remain largely opaque, although what scholars have uncovered is that most policies attempt to balance stakeholder expectations (including users, consumers, advertisers, shareholders, the general public), commercial business goals, and jurisdictional norms and legal demands (which are generally governed by liberal-democratic notions of “free speech”) (Gillespie, 2018; Klonick, 2018; Roberts, 2016). As Roberts (p. 9) summarises, “Internal policies regarding permissible content therefore serve these purposes, first and foremost, rather than responding to, say, social justice or advocacy-related goals.”² Typically, content moderations processes combine company resources (in the form of internal and external content moderators; or technical solutions like automated filters) with user labour (e.g., via flagging or other online reporting mechanisms). These processes can be reactive or proactive (sought after); manual or automated.³ Some platform operators source experts or trained specialists in suicide, human trafficking, child exploitation, domestic violence, terrorism, while others conduct semi-regular audit reviews of moderator decisions to ensure consistency and adherence to policy guidelines.

The most common ‘workable solution’ presented as it relates to content moderation are processes that combine technical and social (human) responses. The manual policing and removal of online hate speech has a long history on news websites, where community managers, employees and journalists have taken up for years as part of their public interest mandate (Pöyhtäri, 2014). Flagging and removal by users has also been a demonstrably effective, albeit labour-intensive, way of having nefarious and offensive speech removed. On the technical side, however, advances in semi- or fully automated systems, including deep learning, show increased promise in identifying inappropriate content and drastically reducing the number of messages human moderators then need to review (Binn et al, 2017; Delort, Arunasalam & Paris, 2011). Tested technical solutions include Reddit’s 2015 ban on two subreddits rife with hate speech (r/fatpeoplehate and r/CoonTown) (Chandrasekharan et al.’s (2017). The study found Reddit’s use of an automatic keyword identification tool effectively encouraged hate-speech accounts to discontinue their use of the site, while remaining accounts reduced their hate speech use by 80 percent. Elsewhere, Galán-García et al. (2014) tested their machine-learning algorithms have been used to track cyberbullies trolling their peers at a Spanish school, effectively narrowing the perpetrators down to three students who then confessed to the bullying. However, these types of tracking techniques raise a number of concerns around algorithmic sorting and institutional surveillance, particularly in educational settings.

Other ongoing research aims to advance a more holistic approach that semi-automates content moderation via more transparent classification systems that try to account for context while providing moderators and users a reason for their classification results (e.g., providing users with an explanation for content deletion) (Risch, J., & Krestel, 2018). Researchers have also found text-based mining alone to be insufficient, and have thus turned towards non-text features like user characteristics as potential datasets for detecting incivility online. In one study, combining certain user features with textual features slightly improved the performance of automated classification results in hate speech detection models (Unsvåg, 2018; Unsvåg & Gambäck, (2018). This tactic again, however, functions on the submission of users to more surveillance.

Empirical researchers are increasingly willing to admit, however, that neither automated nor manual classifications systems can ever be “neutral” or free from human bias: “There can be no formula determining the extent to which different viewpoints need to be

2. Klonick’s historical and qualitative research with insiders tied to Facebook, YouTube and Twitter found finds that platforms developed their moderation systems in accordance with (1) an underlying belief in American constitutional free speech norms; (2) a sense of corporate responsibility (tied mainly to enforcing democratic norms); and (3) the necessity of meeting users’ norms for economic viability (which dominates any sense of corporate responsibility). This is echoed by scholars like Sarah T Roberts, who found content moderation processes at large social media firms are governed by policies that similarly try to balance (1) attracting user-participants and advertisers; 2) responding to jurisdictional norms and legal demands, and 3) remaining profitable and appealing to shareholders.

3. The dominant approach to moderation is based on a human-generated rules-based approach (e.g., white lists; black lists) that are manually processed and resource-consuming; they are not only prone to human bias but often produce erroneous results (Gillespie, 2018; Delort, Arunasalam & Paris, 2011). These automated process are also “difficult to maintain as language, norms, and gaming strategies change” (Binns, Veale, Van Kleek, Shadbolt).

reflected in order for a classifier to be deemed fair” (Binns et al, 2017, p. 411; see also Gillespie, 2016). Human and/or automated content moderation is unlikely to achieve “civil discourse,” a “sanitized” internet or other speech and engagement goals through moderation alone. Therefore, the combination of automated classification and deletion systems and human efforts remains the most effective content moderation strategy currently on offer. Even still, the complex technical and cultural issues that undergird effective moderation have not stopped claims that “more moderation: is the ‘answer’ to combatting the internet’s misinformation campaigns and speech problems. In the few places where they exist (German Network Enforcement Act, 2017), government regulations on private intermediaries’ moderation practices have not been empirically tested for their efficacy or effectiveness.

DISTRUST/ DISSATISFACTION WITH DEMOCRACY

Increasing trust in government institutions (and democracy broadly) is covered by a range of literature on direct and participatory democratic engagement/processes, e-government, and open government.

The creation, selection and use of online platforms that afford citizen participation and deliberation can also enhance government trust (OECD, 2017; Valtysson, 2013). On the one hand, governments need to offer opportunities for citizens to engage in all levels of policy development, including their “design, implementation, monitoring and evaluation” (OECD, 2017, p. 118); such initiatives have the potential to engage and empower all citizens, while “forging a new relationship between cities and local residents, and parliamentarians and citizens” (Simon et al, 2017, p. 11). However, the systems and platforms used in these deliberative processes must also enable these practices to emerge. As mentioned, the techno-social affordances inherent to different online platforms affect and shape the nature of engagement, deliberation and discussions (Bucher & Helmond, 2018; Valtysson, 2013). For example, a recent case study on the use of social media to ‘crowdsource’ the re-writing of the Icelandic Constitution provides empirical support for how social media generates different user responses and modes of engagement. An analysis of this process found that the interactive and participatory affordances of varying social media enabled Council members to provide practical information with constituents, disseminate information to mainstream media and to maintain a positive rapport with citizens via enthusiastic responses to their comments, discussion and suggestions. Yet the study also found that social media – and the government’s use for deliberative democracy – fell short as a space to effectively achieve rational consensus. Social media might *afford* state actors to constitute and engage networked publics, but the practice of extending deliberation to activate the public as a decision-making body requires using government to encourage platform use in this way.⁴

A meta-analysis of empirical research found that contributions to democratic modes of participation differ according to type of democratic innovation: e.g., deliberative forums and surveys better promote exchange of arguments, “whereas referendums and participatory policy making projects are better at giving citizens influence on policy making and involving more people” (Michels, 2011, p. 275). Indeed, international research has found that engaging citizens in deliberative processes often results in profound changes in deliberating citizens’ “frequently in the direction of more common good-

4. The author concedes social media can engender strong, effective and civil political *discussion*, but operationalises ‘deliberation’ by the Habermasian standard of achieving rational consensus – the point at which social media falls short.

oriented policies” (Bächtiger et al., 2011, p. 5). This finding seems to stand even in the context of highly polarised issues, although the influence of deliberation on preference transformation towards less simplistic measures is also thought to be moderated by a period of internal reflection (Bächtiger et al., 2011). This period of “deliberation within” may in fact be more important to preference or opinion change than discussion itself (Goodwin, 2003), suggesting that time to adequately reflect upon information provided before engaging in deliberative discussion is more important conversation itself. Empirical research supports that preceding online modes of deliberation with a period of internal reflection is effective, yet quite the challenge in a communicative environment premised on immediacy, constant updates and a dynamic information cycle. It’s worth noting as well that ‘deliberation’ in these research contexts are typically operationalised as engaged, civil discussion, which precludes Habermas’ requisite of achieving rational consensus.

Transparency, access and design are essential to generating positive results from digital government initiatives, and when done effectively have been shown to increase positive feelings and citizen trust in local government (Leininger, 2015; Kern, 2017; Swaner, 2017). The shift towards “e-government,” whereby information and communication technologies (ICTs) are used to improve public sector activities, aims to improve state services and engage more citizens in democratic processes (eGovernment for Development, 2008; OECD, 2017). Similarly, ‘open government’ initiatives are designed to increase transparency, inclusiveness and thus institutional trust, but have largely advocated for “open data” initiatives that make government data freely available for public use – a insufficient solution on its own (Lourenço, 2015). Instead, adequate design, resourcing and accessibility are central to successful e-government and open government initiatives. Some evidence supports the finding that governments that have created usable, intelligible websites, and offer non-exclusionary solutions for those lacking computer and internet access or basic digital literacy skills, have been most successful in their e-government initiatives and constituent satisfaction (UN, 2018).

Finally, civics education, or educating children in schools on “good citizenship” has been positively associated with increased political engagement, particularly for socially and economically marginalised groups (Neundorf, Niemi & Smets, 2016; Van de Werfhorst, 2017).

PART 2.

COMMON RECOMMENDATIONS ACROSS THE LITERATURE:

This research confirms that there is, at present, a troubling dearth of scientific, empirical, evidence-based scholarship that tests or aims to validate “workable solutions” to the seven key threats to democracy we’ve identified in this project. A significant majority of the academic research relating to these seven problems is *normative* in approach; meaning, it presents theoretically sound arguments about the way things “ought to be” if democracy is to be “reclaimed” from incivility and a rogue form of capitalism in the digital age.

While there is certainly a place in both academia and policy for normative (and critical) research, there is also a need to begin testing many of the propositions and recommendations being made and instituted by industry, governments, NGOs and civil society to not only measure and extend what’s working to other places or contexts, but so that future normative prescriptions are informed by evidence beyond the anecdotal (or budgeting restrictions), as well. We need to start creating testable knowledge beyond good ideas and theory-informed ‘recommendations’ (although again, there is a place for that).

What follows is a summary of prescribed recommendations that appeared most frequently across the literature as advocated by scholars, legal experts, advocacy groups and other civil society organisations. It is not an exhaustive list. It is also worth noting that many of these recommendations emerge from – and also recommend – a multi-stakeholder approach to internet governance that represents and balances the interests of industry, state, and civil society. In such cases we must consider carefully how people from civil society are included in a way that take account of the power imbalances between this group and the other two.

1. POLICY / LEGAL SOLUTIONS

Proposals here range from calls to review and overhaul existing legislation; create new legislation; institute new oversight bodies or inter-government agencies (e.g., Establishing an independent Commissioner and/or oversight board with the power to impose fines on private intermediaries; remove social media from ISPs; regulate/vet algorithms to minimise injury to public interest); or to improve regulations on content moderation.

In the US context, many scholars are calling for regulators to revisit and consider overhauling the governing legislation around moderation issues (Gillespie, 2018; Klonick, 2018; Laidlaw, 2015). This legal framework is embedded with distinctly American normative values around free speech and free enterprise that have significant ramifications for the rest of the world (particularly places that do not share these same cultural values). Currently, the governing legislation for social media content is Sec 230 of the 1996 Communications Decency Act, which absolves all internet companies and intermediaries – from ISPs to platforms to content makers – from liability for any content they host while allowing them to also delete any content they wish without consequence. By this law, platforms are not legally obliged to remove or filter content that does not violate the company's terms of service, even in cases where legal courts find such content defamatory (e.g. Hassell vs Bird, 2018). Conversely the rule simultaneously allows platforms to remove or filter content that constitutes 'free speech' or some other constitutional protection without consequence. Revising this regulation, however, first requires determining whether platforms constitute media companies (broadcasters or editors?); state actors; a "public" or public town; a (public) utility or some hybrid formation for which current speech regulations do not yet exist or apply. How private intermediaries are defined by law thus inform their regulation, yet these are questions that have only begun being asked let alone resolved.⁵

2. MORE CORPORATE TRANSPARENCY

Currently the lack of transparency around moderation practices presents challenges to accountability, governance, and the ability to apply public and legal pressure. Effective examination of the "moderation apparatus" includes access to private intermediaries' a) content moderation policies; b) the sociotechnical mechanisms used in their enforcement (e.g., system design, labour, organisational culture); c) business expectations the apparatus must serve; d) the justifications articulated to support these criteria (Gillespie, p. 12-13). Expanding empirical research to improve moderation processes requires private intermediaries make these processes and practices accessible to researchers.⁶ Access to formal documents or policies related to moderation processes are also generally unavailable, and in cases where semi- or fully-automated techniques contribute to content moderation, little insight is available about how these processes work, their efficacy or the algorithms behind them.

5. For example, Klonick (2018) advises against treating platforms as state actors bound to First Amendment (free speech) law given the difficulties of determining "when a private party's behaviour constitutes state action," e.g., in what situations or context private property functions as a public space (p. 1659). Regulating as media companies is also a tenuous proposition as in the US, the regulation of broadcasters is based on redressing matters arising from spectrum scarcity and "invasive" nature that don't apply in the same way to the digital context (e.g., the lack of spectrum space justifies need to ensure media meet public interest needs, although it's certainly possibly that the right to platform access might challenge future scarcity claims). Treating platforms as 'forums' or town squares would give them their own First Amendment speech rights, which invests platforms with what some believe is too much independent regulatory power.

6. Research has generally found that the state of platform content moderation by private intermediaries is a closed, fractured and private process inaccessible to users, the public and even researchers; these conditions at least partly account for the lack of empirically-based research on content moderation practices. Despite calls for more transparency, content moderation remains a closed process, if not protected as an industrial trade secret. Human content moderators are themselves sequestered into silence through non-disclosure agreements (NDAs), fragmented and siloed as their own internal department (separate from engineers or company workforce) if not outsourced to private "boutique" firms, overseas micro-labor sites or call centres in the form of low-waged, low-status labor. In the latter case, moderation jobs are often cast as "customer-service" [managed and staffed by women](#), which stands apart from the higher-status, higher-paid, more powerful sectors of engineering and finance, which are overwhelmingly male.

3. BETTER DESIGN

The consideration of platform affordances shows how design can influence the way individuals, organizations and institutions make decisions around platform uses/objectives. Pro-social and democratic values must be encoded into the infrastructure of the internet, including algorithms (Annanny, 2016; O’Neil, 2016). Assessing what normative values are being built into an ever-evolving internet infrastructure, whose values they represent and what forms of social life they afford are primary research questions guiding this approach. At present, the normative values embedded into these global private intermediaries – e.g., openness, connectedness, free speech, etc. – are not culture-neutral norms. While this creates a range of problems for an internet governed by global private intermediaries (including wholesale censorship regimes like China and North Korea’s banning of Facebook and Twitter), it is also the first step towards designing more deliberative spaces, pro-social tools and online environments.

4. IMPROVE CONTENT MODERATION

Calls range from the standardisation of industry-wide “best practices” (e.g., Harvard’s Berkman Center for Internet and Society, 2019) to more transparency and researcher access (Gillespie, 2018; Roberts, 2016). Suggestions to the former recommendation would require: a) Corporate transparency, consistency, clarity, and clear mechanisms for customer recourse around moderation complaints; b) Corporate grievance mechanisms that are transparent, accessible and in accordance with international human rights law; c) A multi-stakeholder governance approach that demands platforms engage in public dialogue with relevant organisations (Anti-Defamation Leagues, Digital Rights Foundation, National Network to End Domestic Violence, etc). (Gillespie, 2018; Harvard’s Berkman Center for Internet and Society, 2019). Suggestions also include making content moderation an organisational priority rather than department silo, and one that brings together legal, customer service, security, privacy, safety, marketing, branding, and personnel to create a unified approach to resolving such a complex issue (Gillespie, 2018; Klonick, 2018; Roberts, 2016).

PART 3.

ACTIVITIES IN

NEW ZEALAND

Given that in our literature review we identified very little empirically tested evidence of any quality, and a plethora of expert opinion, we also looked to identify activities relevant to digital democracy in New Zealand. We did not do so systematically, and we have little or no evidence as to their impact on optimising opportunities, reducing threats, or improving democratic participation. This section therefore is purely a high level scan of local activities.

1. Two 2016 reports (Open Data NZ, 2016a, 2016b), describe public engagement measures that have been undertaken in New Zealand. For example, The Open Data Charter Public Consultation “ran throughout August and September 2016...designed to give interested and potentially affected parties with the opportunity to provide their thoughts about whether New Zealand should adopt the Open Data Charter and the impacts if it is was to be adopted” (Open Data NZ, 2016b, p. n.p.). Further, a nationwide survey was conducted, “asking people what the top 10 datasets they would like to see released were”, to which they received “over 400 responses” (n.p.).
2. We have previously discussed two Loomio experiments conducted by Wellington City Council to consult on an alcohol management strategy, and Statistics New Zealand on gender categorization.
3. Work is currently being undertaken by Government Information Services (2018) in the area of e-government inclusivity. For example a Digital Inclusion Ministerial Advisory Group had been established (InternetNZ 2018). and the Government Online Engagement Service (GOES) was established in 2011, which has been piloting a survey tool asking people “how they would like to have their say with government” (Government Information Services, 2018, p. 11). Results suggest a thirst in New Zealand for initiatives in digital participative democracy.
4. The New Zealand government developed a cyber security strategy in 2012 (updated in 2016), and established the National Cyber Policy Office in the same year, and Netsafe in 2016, under the Harmful Digital Communications Act. (InternetNZ, 2017a).
5. A review of the Privacy Act by the Privacy Commissioner, which includes a focus on data protection for consumers (Privacy Commissioner, 2016);
6. The monitoring of the Harmful Digital Communications Act by the Human Rights Commission, in order to ensure we “strike the right balance between freedom of expression and our need as a community to challenge our bullying culture and protect people who are under attack” (Human Rights Commission, 2015).
7. A submission to the United Nations committee that oversees the Convention on the Elimination of Racial Discrimination, in order to improve the Human Rights Act to better account for online hate speech (Human Rights Commission, 2017).
8. The Local Government Excellence Programme (LGNZ, 2016) is a program which councils can choose to sign up to voluntarily, and provides them with an independent assessment “across four priority areas that the general public and business communities have told us are important to them” (p. 3). Councils are then ranked on their performance. This programme was specifically designed to counter declining trust in local councils. While the programme was only established in 2016, the 2017 Local Government Survey (LGNZ, 2017) showed a slight improvement in overall satisfaction.
9. Design+Democracy and RockEnrol, academic and civil society organisations, have undertaken work in the area of youth engagement. Design+Democracy’s (2017) On the Fence describes itself as a “gameful questionnaire” and a “fun educational tool that helps young undecided and first-time voters engage directly with issues by matching their personal values with political candidates and parties”. In the six week period before the 2017 election day, the site received “170,000 unique visitors”, who, on average, “achieved a 92% completion rate”. “A Horizon Research survey showed that the site encouraged 30,000 non-voting 18–34 year olds to vote, representing 7% of the total eligible youth population”.

CONCLUSIONS

While our literature review was not exhaustive, the general finding was that there is a sheer dearth of empirical evidence in this area. This lack of evidence of any quality leads us to conclude that there is a critical need for people in government, civil society, NGOs and enterprise to invest in researchers and projects who will do pre- and post-testing of the solutions that stakeholders are recommending, including any recommendations the government decides to take up. It is critical that people in the New Zealand Government see whether or not what's being done / put in place is working. New Zealand would break significant ground in that regard.

Undertaking research in the following areas is critical:

- Understanding and agreeing what social media platforms constitute. Are they media companies (broadcasters or editors?), telecommunication companies, state actors, a “public” or public square, a utility, or some hybrid of the above? And given this, what policy and legal solutions work best to minimise public harm?
- Exploring the impact of public investment in new competitive solutions to forge an alternative digital future and facilitate a more democratic internet.
- Determining whether the creation and support of new modes of collective action can achieve broader structural changes across the political economy and culture.
- Understanding how moderation processes and practices are used by social media companies, how they are implemented, the algorithms behind them and the impacts.
- Assessing the impact of participatory decision making processes at a central and local government level using platforms with pro-social and democratic values encoded within them.
- What is the impact of standardised, industry wide best practice content moderation.
- Can public attitude campaigns, and associated collective action, create a shift in social media companies practices to rebalance power between civil society and private social media companies?

It is critical also that as multi stakeholder recommendations are sought that people in government and civil society mitigate the risk that experts reproduce “solutions” that fit the professional discourses in which they’re embedded. To do this it is important that people in government ask multi-stakeholder group participants:

1. What if any evidence they have for the suggestions made?
2. What experiences inform these recommendations and why do they identify them as workable solutions over others?
3. How do they imagine testing their effectiveness?

It will be critical to make transparent and visible the values, experiences, and outcomes that are underlying recommendations made when there is a lack of evidence available.

REFERENCES

- Ananny, M. (2016). Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness. *Science, Technology, & Human Values*, 41(1), 93-117. doi:10.1177/0162243915606523
- Bächtiger, A., Steenbergen, M., Gautschi, T., & Pedrini, S. (2011). Deliberation in Swiss direct democracy: A field experiment on the expulsion initiative. *NCCR Newsletter*, 8.
- Bakker, P. 2010. "Journalistiek zonder infrastructuur – een reële optie? (Journalism without Infrastructure- a Realistic Option?)." *Tijdschrift voor Communicatiewetenschap* 38 (3), 250–258.
- Betkier, M. (2018). Moving beyond consent in data privacy law. An effective privacy management system for Internet services. Victoria University, Wellington.
- Binns, R., Veale, M., Van Kleek, M., & Shadbolt, N. (2017). Like trainer, like bot? Inheritance of bias in algorithmic content moderation. *International Conference on Social Informatics* (pp. 405-415). Springer, Cham.
- Bodkin-Andrews, G., Newey, K., O'Rourke, V., & Craven, R. (2013). Promoting resiliency to counter racism: the lived wisdom within Aboriginal voices. *Psychological perspectives on racism*. In *Psych: The Bulletin of the Australian Psychological Society Ltd*, 35(4), 14.
- Chandrasekharan, E., Pavalanathan, U., Srinivasan, A., Glynn, A., Eisenstein, J., & Gilbert, E. (2017). You can't stay here: The efficacy of reddit's 2015 ban examined through hate speech. *Proceedings of the ACM on Human-Computer Interaction*, 1(CSCW), 31.
- Delort, J.Y., Arunasalam, B & Paris, C. (2011). Automatic Moderation of Online Discussion Sites. *International Journal of Electronic Commerce*, 15(3), 9-30. doi: 10.2753/JEC1086-4415150302
- Design+Democracy. (2017). On the Fence. Massey University. Retrieved from <http://designdemocracy.ac.nz/initiatives/on-the-fence>
- Diakopoulos, N., & Naaman, M. (2011, March). Towards quality discourse in online news comments. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work* (pp. 133-142). ACM.
- Dubow, T. (2017). Civic Engagement: How Can Digital Technologies Underpin Citizen-Powered Democracy? Retrieved from https://www.rand.org/pubs/conf_proceedings/CF373.html
- eGovernment for Development. (2008). What is eGovernment? Retrieved from <http://www.egov4dev.org/success/definitions.shtml>

- Flew, T., Martin, F., & Suzor, N. P. (2019). Internet regulation as media policy: Rethinking the question of digital communication platform governance. *Journal Of Digital Media And Policy*, 10(1), 33-50.
- Fuchs, C., & Trottier, D. (2017). Internet surveillance after Snowden: A critical empirical study of computer experts' attitudes on commercial and state surveillance of the Internet and social media post-Edward Snowden. *Journal of Information, Communication and Ethics in Society*, 15(4), 412-444. doi:10.1108/JICES-01-2016-0004
- Galán-García, P., de la Puerta, J. G., Gómez, C. L., Santos, I., & Bringas, P. G. (2014, 2014//). Supervised Machine Learning for the Detection of Troll Profiles in Twitter Social Network: Application to a Real Case of Cyberbullying. Paper presented at the International Joint Conference SOCO'13-CISIS'13-ICEUTE'13, Cham.
- Gargliardone, I., Gal, D., Alves, T., & Martinez, G. (2015). "Countering Online Hate Speech." 13-14. Unesco Publishing, <http://www.scribd.com/doc/269217679/UNESCO-Countering-Online-Hate-Speech#scribd>
- Gillespie, T. (2018). *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.
- Goodwin, I. (2003). "Power to the people? Web 2.0, Facebook, and DIY cultural citizenship in Aotearoa New Zealand." *MEDIANZ: Media Studies Journal of Aotearoa New Zealand* 12.2 (2014): 110-134.
- Government Information Services. (2018). How digital can support participation in government. Retrieved from Wellington: <https://www.digital.govt.nz/standards-and-guidance/engagement/online-engagement/research-how-digital-can-support-participation-in-government/>
- Groshek, J., & Cutino, C. (2016). Meaner on mobile: Incivility and impoliteness in communicating contentious politics on sociotechnical networks. *Social Media+Society*, 2(4). doi: 10.1177/2056305116677137
- Halpern, D., & Gibbs, J. (2013). Social media as a catalyst for online deliberation? Exploring the affordances of Facebook and YouTube for political expression. *Computers in Human Behaviour*, 29(3), 1159–1168. doi: 10.1016/j.chb.2012.10.008
- Hassell v. Bird (California Court of Appeal 2016).
- Hicks, T. (2018). Defending and Recovering American Election Systems. *Brown Journal of World Affairs*, 24(2), 97-108.
- Hille, S. & Bakker, P. (2014) Engaging the Social News User. *Journalism Practice*, 8(5), 563-572. doi: 10.1080/17512786.2014.899758
- Hmielowski, J. D., Hutchens, M. J., & Cicchirillo, V. J. (2014). Living in an age of online incivility: Examining the conditional indirect effects of online discussion on political flaming. *Information, Communication & Society*, 17(10), 1196–1211. doi:10.1080/1369118X.2014.899609
- Human Rights Commission. (2015). HRC to monitor Harmful Digital Communications Act. Retrieved from <https://www.hrc.co.nz/news/hrc-monitor-harmful-digital-communications-act/>

- Hutchby, I. (2001). *Conversation and technology : From the telephone to the Internet*. Cambridge, UK : Oxford, UK ; Malden. Mass.: Polity Press ; Blackwell.
- Internet Governance Forum. (2015). Recommendations on terms of service and human rights. Retrieved from <https://www.intgovforum.org/cms/documents/igf-meeting/igf-2016/830-dcpr-2015-output-document-1/file>
- InternetNZ. (2017a). State of the Internet 2017. Retrieved from <https://internetnz.nz/sites/default/files/SOTI%20FINAL.pdf>
- InternetNZ. (2018). Solving Digital Divides Together. Retrieved from <https://internetnz.nz/publications/solving-digital-divides-together-0>
- Jackson, S. K., & Kuehn, K. M. (2016). Open Source, Social Activism and “ Necessary Trade-offs” in the Digital Enclosure: A Case Study of Platform Co-operative, Loomio. *org. tripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society*, 14(2), 413–427-413–427.
- Jakubowicz, A., Dunn, K., Mason, G., Paradies, Y., Bliuc, A.-M., Bahfen, N., & Connelly, K. (2017). *Cyber Racism and Community Resilience: Strategies for Combating Online Race Hate*. Springer.
- Kamara, I. (2017). Co-regulation in EU personal data protection: the case of technical standards and the privacy by design standardisation ‘mandate’. *European Journal of Law and Technology*, 8(1).
- Kavada, A. (2012). Engagement, bonding, and identity across multiple platforms: Avaaz on Facebook, YouTube, and MySpace. *MedieKultur: Journal of Media and Communication Research*, 28(52).
- Kern, A. (2017). The effect of direct democratic participation on citizens’ political attitudes in Switzerland: The difference between availability and use. *Politics and Governance*, 5(2), 16-26. doi:<http://dx.doi.org/10.17645/pag.v5i2.820>
- Klonick, Kate. (2016). Re-shaming the debate: Social norms, shame and regulation in an Internet age. *Maryland Law Review*, 75(4), 1029-1065.
- Klonick, K. (2018). The New Governors: The People, Rules, and Processes Governing Online Speech. *Harvard Law Review*, 131(6), 598-670.
- Laidlaw, E. (2015). *Regulating Speech in Cyberspace Gatekeepers: Human Rights and Corporate Responsibility*. Emily B. Laidlaw.
- Lanois, P. (2016). Data Protection in Singapore: What Have We Learned? *Journal of Internet Law*, 20(2), 1-16.
- Leininger, A. (2015). Direct Democracy in Europe: Potentials and Pitfalls. *Global Policy*, 6(S1), 17-27. doi:10.1111/1758-5899.12224
- LGNZ. (2016). Local Government Excellence Programme prospectus –overview. Retrieved from <http://www.lgnz.co.nz/assets/Uploads/Final-Local-Government-Excellence-Programme-PAF-10.2.17.pdf>

- Lourenço, R. P. (2015). An analysis of open government portals: A perspective of transparency for accountability. *Government Information Quarterly*, 32(3), 323-332. doi:<https://doi.org/10.1016/j.giq.2015.05.006>
- Marda, V., & Milan, S. (2018). *Wisdom of the Crowd: Multistakeholder Perspectives on the Fake News Debate*. Internet Policy Review series. Annenberg School of Communication.
- Michels, A. (2011). Innovations in democratic governance: how does citizen participation contribute to a better democracy? *International Review of Administrative Sciences*, 77(2), 275-293. doi:10.1177/0020852311399851
- Narayanan, A., & Reisman, D. (2017). The Princeton Web Transparency and Accountability Project. In T. Cerquitelli, D. Quercia, & F. Pasquale (Eds.), *Transparent data mining for big and small data*. New York: Springer.
- Network Enforcement Act (NetzDG) (2017). German Bundestag.
- Neundorf, A., Niemi, R. G., & Smets, K. (2016). The compensation effect of civic education on political engagement: How civics classes make up for missing parental socialization. *Political Behavior*, 38(4), 921-949.
- Nsangi, A., Semakula, D., Oxman, A. D., Austvoll-Dahlgren, A., Oxman, M., Rosenbaum, S., ... & Chalmers, I. (2017). Effects of the Informed Health Choices primary school intervention on the ability of children in Uganda to assess the reliability of claims about treatment effects: a cluster-randomised controlled trial. *The Lancet*, 390(10092), 374-388.
- O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. New York: Crown.
- Oboler, A., & Connelly, K. (2018). Building SMARTER Communities of resistance and solidarity. *Cosmopolitan Civil Societies: An Interdisciplinary Journal*, 10(2), 99.
- OECD. (2017). *Trust and public policy: How better governance can help rebuild public trust*. Paris: OECD Publishing.
- Ombler, J., Russell, M., & Rivera-Munoz, G. (2016). Local councils and public consultation: extending the reach of democracy. *Policy Quarterly*, 12(4).
- Open Data NZ. (2016b). Overview of Open Data NZ's Work in 2016. Retrieved from <https://www.data.govt.nz/assets/Uploads/Report-Open-Data-NZ-2016-Overview2.pdf>
- OSCE. (2017). Joint declaration on freedom of expression and "fake news", disinformation and propaganda. Retrieved from <https://www.osce.org/fom/302796>
- Papacharissi, Z. (2004). Democracy online: Civility, politeness, and the democratic potential of online political discussion groups. *New Media & Society*, 6(2), 259-283.
- Pöyhtäri, R. (2014). Limits of hate speech and freedom of speech on moderated news websites in Finland, Sweden, the Netherlands and the UK. *Annales-Series historia et sociologia izhaja štirikrat letno*, 24(3), 513-522.
- Privacy Commissioner. (2016). Privacy Commissioner's Review of the Privacy Act. Retrieved from <https://www.privacy.org.nz/the-privacy-act-and-codes/privacy-law-reform/privacy-commissioner-s-review-of-the-privacy-act/>

- Ray, R., Brown, M., Fraistat, N., & Summers, E. (2017). Ferguson and the death of Michael Brown on Twitter: #BlackLivesMatter, #TCOT, and the evolution of collective identities. *Ethnic and Racial Studies*, 40(11), 1797-1813. doi:10.1080/01419870.2017.1335422
- Risch, J., & Krestel, R. (2018). Delete or not delete? semi-automatic comment moderation for the newsroom. In *Proceedings of the First Workshop on Trolling, Aggression and Cyberbullying (TRAC-2018)*, 166-176.
- Roberts, S. T. (2016). Commercial content moderation: digital laborers' dirty work. *Media Studies Publications*. 12. <https://ir.lib.uwo.ca/commpub/12>
- Rowe, I. (2015). Civility 2.0: A comparative analysis of incivility in online political discussion. *Information, Communication & Society*, 18(2), 121-138.
- Rushkoff, D. (2014). "Loomio: the Occupy inspired app for consensus decision making." Retrieved from <http://www.shareable.net/blog/loomio-the-occupy-inspired-app-for-consensus-decision-making>
- Sandoval, M. (2016). Fighting Precarity with Co-Operation? Worker Co-Operatives in the Cultural Sector. *New Formations*, 88, 51-68. doi: 10.3898/newf.88.04.2016
- Santana, Arthur D. (2014). Virtuous or Vitriolic. The Effect of Anonymity on Civility in Online Newspaper Reader Comment Boards. *Journalism Practice*, 8 (1), 18-33.
- Scholz, T. (2016). Platform Cooperativism: Challenging the Corporate Sharing Economy. New York: Rosa Luxemburg Stiftung.
- Shoorbajee, Z. (2018, February 26). New EAC chairman will continue to focus on election security. *Cyberscoop*. Retrieved from <https://www.cyberscoop.com/thomas-hicks-eac-chair/>
- Simon, J., Bass, T., Boelman, V., & Mulgan, G. (2017). Digital democracy: the tools transforming political engagement. Retrieved from <https://www.nesta.org.uk/report/digital-democracy-the-tools-transforming-political-engagement/>
- Srnicek, N. (2017). *Platform capitalism*. Cambridge, UK: Polity Press.
- Stats NZ. (2018, April 10). New Zealand Government. Retrieved from <https://www.stats.govt.nz/reports/sex-gender-and-sexual-orientation>
- Stohl, C., Stohl, M., & Ganesh, S. (2018). Digital media and human rights: Loomio, Statistics New Zealand, and gender identity. In A. Brysk & M. Stohl (Eds.), *Contracting Human Rights: Crisis, Accountability, and Opportunity*, 232-249. Cheltenham, UK: Edward Elgar Publishing.
- Swaner, R. (2017). Trust Matters: Enhancing Government Legitimacy through Participatory Budgeting. *New Political Science*, 39(1), 95-108.
- Sweet, M., Pearson, L., & Dudgeon, P. (2013). @Indigenousx: A Case Study of Community-Led Innovation in Digital Media. *Media International Australia*, 149(1), 104-111. doi:10.1177/1329878x1314900112
- U.S. Election Assistance Commission. (2016). Election Administration and Voting Survey. Retrieved from https://www.eac.gov/assets/1/6/2016_EAVS_Comprehensive_Report.pdf

- UN. (2018). United Nations E-Government Survey 2018. Retrieved from New York: https://publicadministration.un.org/egovkb/Portals/egovkb/Documents/un/2018-Survey/E-Government%20Survey%202018_FINAL%20for%20web.pdf
- Unsvåg, E. F. (2018). Investigating the Effects of User Features in Hate Speech Detection on Twitter (Master's thesis, NTNU).
- Unsvåg, E. F., & Gambäck, B. (2018). The Effects of User Features on Twitter Hate Speech Detection. In Proceedings of the 2nd Workshop on Abusive Language Online (ALW2), 75-85.
- Valtysson, B. (2013). Democracy in disguise: the use of social media in reviewing the Icelandic Constitution. *Media, Culture & Society*, 36(1), 52-68. doi:10.1177/0163443713507814
- Van de Werfhorst, H. G. (2017). Vocational and academic education and political engagement: The importance of the educational institutional structure. *Comparative Education Review*, 61(1).
- Zuboff, S. (2019). Surveillance Capitalism and the Challenge of Collective Action. *New Labor Forum*, 28(1), 10-29.

APPENDIX 3: REPORT ON QUANTITATIVE SURVEY

D I G I T A L T H R E A T S T O D E M O C R A C Y

R E P O R T
O N
Q U A N T I T A T I V E
S U R V E Y

S O C I A L M E D I A
A N D P O L I T I C A L

E N G A G E M E N T I N N E W Z E A L A N D



Luminate

D I G I T A L T H R E A T S T O D E M O C R A C Y

This report is part of the Digital Threats to
Democracy research project.

To see the rest of the reports and the
overall findings go to digitaldemocracy.nz

ISBN: 978-0-473-48026-4

EXECUTIVE SUMMARY

As part of this research into digital democracy in New Zealand, we commissioned a representative survey of 1000 New Zealanders by UMR in September and October 2018. The survey was designed to elicit the views and experiences of people using social media and digital platforms relevant to democracy (e.g. participating in debates about issues of public policy on social media.) Some key findings from the analysis that have emerged:

- Social media platforms are used for political activity by minority ethnic groups more than Pākehā. Indicating their potential use as a tool for engagement in formal democratic system.
- Stated trust in news online may be low, but perceptions of information credibility are driven by trust in friends, family and organisations.
- Most people still rely on mainstream media for information about a key political issue (decriminalisation of marijuana) but friends and family and online news feature strongly.
- New Zealanders are relatively accurately informed about the opinions of others with regard to the decriminalisation of marijuana.
- There is evidence that New Zealanders who believe their views are in the minority on decriminalisation of marijuana are less willing to share their views both offline and online. This suggests social media platforms replicate rather than overcome existing barriers to engaging in less formal processes of democracy (public discussion and discourse) for people who hold minority views.

WHO DID WE SURVEY?

UMR research surveyed 1000 New Zealanders aged 18 years and over between the 27th of September and 2nd of October 2018. The data was weighted by region, age, gender identity (including a category of “different”), and ethnicity to accurately represent the New Zealand population. See the appendix for further methodological details.

Use of social media is high: Facebook dominates

Most of those asked (87%) named Facebook as a social media platform they use. A smaller proportion of people named Twitter (24%) and Instagram (21%) as social media they use. A very small proportion identified using YouTube (3%). It is a lower proportion than we would have expected to use YouTube. This was a non-promoted question and it is possible that many people do not define YouTube as a “Social Media” platform. A small number of people, 9%, use no social media at all. Non users were slightly more likely to be men, and people aged over 60.

USE OF SOCIAL
MEDIA AND
ACTIVITIES RELATED
TO DEMOCRACY

Around a quarter of the sample (24%) used social media to engage with ‘political issues’ or politicians. Around 62% of people asked said they did not use social media for activities related to political or social issues.

There was an interesting pattern of engagement reported by minority ethnic groups. Of those Māori, Pacific and Asian people who were in the survey, 28%, 34% and 31% respectively said they ‘Belong to a group on a social media platform that is involved in political or social issues, or that is working to advance a cause’ identified as “different identify”. This compares to 23% of NZ European/Pākehā people in the sample. Similarly greater proportions of Māori (32%), Pacific (43%) and Asian (28%) people ‘Follow any elected officials, candidates for office or other political figures on a social media platform’ than NZ European/Pākehā people (23%).

POLITICAL ACTIVITIES PEOPLE ENGAGE IN ON
SOCIAL MEDIA: MINORITIES ARE MORE ACTIVE

Figure 1 shows that nearly half of people asked liked, signed or shared political or social material including petitions online. Fewer posted their own thoughts or encouraged people to take action or vote.

Have you ever used social media platforms to:

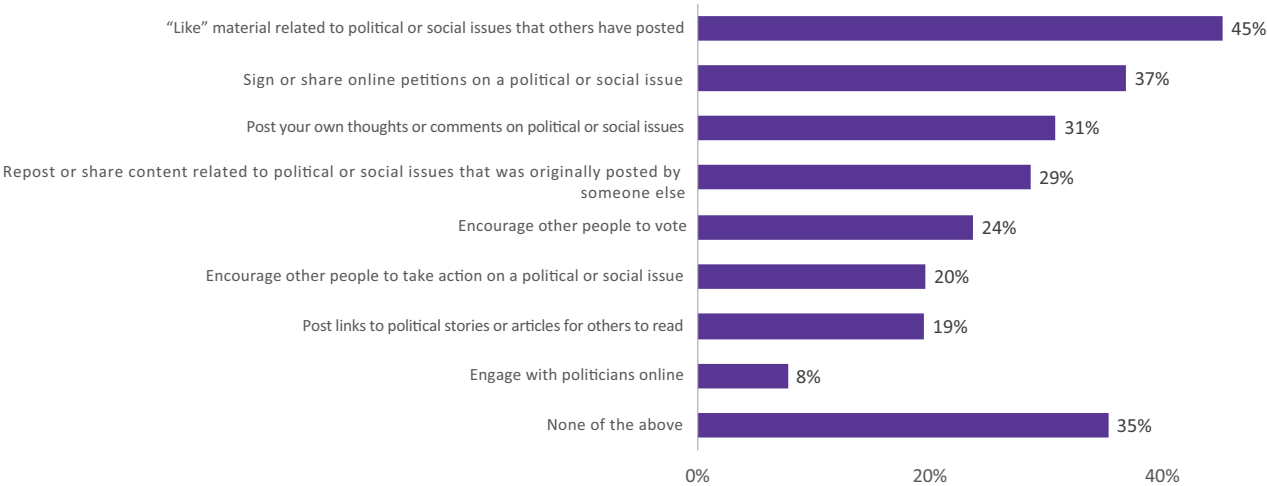


Figure 1: Types and rates of activities related to democracy

Higher proportions of Māori and Pacific people in the survey used social media to encourage people to vote or take action on a political issues than NZ European/Pakeha people (31%, and 37%, compared to 22%). Greater proportions of younger people used social media to encourage voting compared to older people also.

STATED TRUST IN NEWS ONLINE MAY BE LOW, BUT PERCEPTIONS OF INFORMATION CREDIBILITY ARE DRIVEN BY TRUST IN FRIENDS, FAMILY AND ORGANISATIONS.

Expressed trust in online news was low - half of the people surveyed (49%) distrusted news shared online, and 79% agreed they used other sources to check accuracy of online news. However, around 70% of people agreed that they “tend to assess the credibility of online news depending on how much I trust the person who shared it”. When considered in the context of the finding that 88% of survey respondents trust information (more generally) from friends and family, methods of verifying the credibility of online news appear to be highly dependent on social relationships. Existing research is clear that trust in individual people and organisations has a high degree of influence on perceptions of information accuracy, as does perceived expertise (not actual expertise). New Zealanders are, it appears, no different in that when news or information comes via those we trust we reduce our scepticism.

Interestingly trust in print sources of news, as well as TV and Radio was much higher (83% and 82%) than online news sources that were not Facebook or Twitter.

Generally, women showed slightly higher propensity to say they trusted online and other sources of news and information than men.

EXPLORING SPIRALS OF SILENCE AND POLARISATION EFFECTS IN NEW ZEALAND

In this survey we wanted to explore the role of social media in people’s willingness to engage in discussions about political issues. For this we needed a public policy issue that had high levels of media exposure and was moderately but not highly polarising. In a separate survey, we tested two policy issues to assess levels of public awareness and polarising effect. The decriminalisation of marijuana was the best match to the criteria above, and therefore chosen for use in this survey. We were not exploring or reporting people’s attitudes about marijuana policy more generally.

In Figure 2 we see that there is a relatively normal distribution of beliefs about how polarising the issue is in New Zealand.

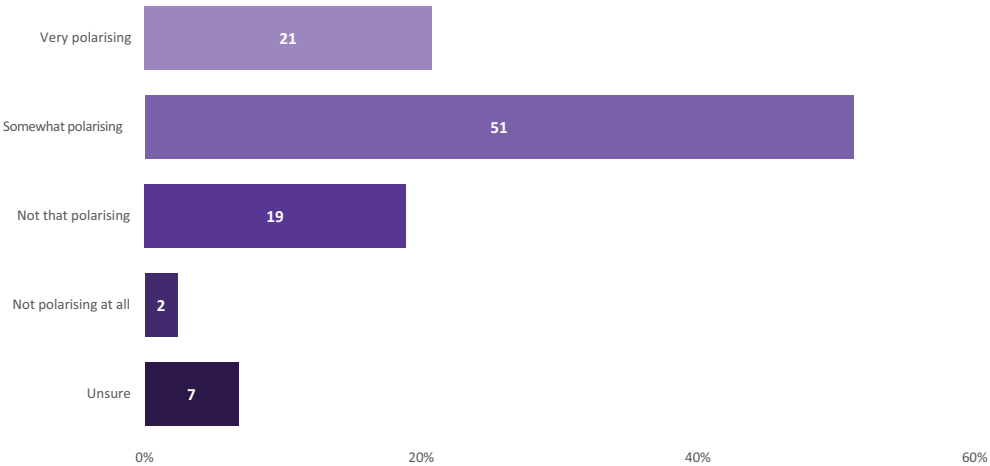


Figure 2: Respondents beliefs about how polarising the marijuana issue is in New Zealand

1. Hampton, K.N., Rainie, L., Lu, W., Dwyer, M., Shin, I., & Purcell, K. (2014). “Social Media and the ‘Spiral of Silence.’ Pew Research Center, Washington, DC. Available at <http://www.pewinternet.org/2014/08/26/social-media-and-the-spiral-of-silence/>

MOST PEOPLE STILL RELY ON MAINSTREAM MEDIA FOR INFORMATION ABOUT A KEY POLITICAL ISSUE (DECRIMINALISATION OF MARIJUANA), BUT FRIENDS AND FAMILY AND ONLINE NEWS FEATURE STRONGLY

When asked if they were interested or knowledgeable about the issue of decriminalisation of marijuana, 64% of respondents said they were interested and 55% knowledgeable about it as an issue. When asked how much “have you read, seen or heard about the debate surrounding decriminalising of marijuana for personal use from TV or Radio?” most people (78%) said they got information about the issue from these mainstream media sources. It was not a finding that differed much based on gender, age or ethnicity or location.

The next most common source of information were friends and family (63%). More younger people in the survey (around 60%) got information from friends and family on the issues than those over 65 (49%). There were also ethnic differences - 77% of Māori and 75% of Pacific people said they got information on the issue from friends and family, compared with 63% and 64% of NZ Europeans/Pākehā and Asian people.

When asked how much “have you read, seen or heard about the debate surrounding decriminalising of marijuana for personal use from online news sources excluding Facebook/Twitter, 61% said they received information from online news sources. The most common sources were Stuff.co.nz and the NZ Herald. More younger people got information online than older people. A majority of people (53%) were still getting information on the issue from print media.

Of those people who used social media, 58% got their information on the issue from Facebook and 50% from Twitter.

AROUND A THIRD OF PEOPLE WHO OPPOSE THE DECRIMINALISATION OF MARIJUANA DO NOT CORRECTLY IDENTIFY THAT THEY ARE IN THE MINORITY, ANOTHER THIRD DO.

Published research suggests that people who hold minority views on some social issues, for example racist views on immigrants and indigenous people (ref), widely overestimate how many people share their views, due to the amplification, polarising, and echo chamber effects of digital and social media. This is called the false consensus effect. Research also shows those with views in the majority, or not polarized can come believe most others hold more extreme views for the same reasons - known as pluralistic ignorance

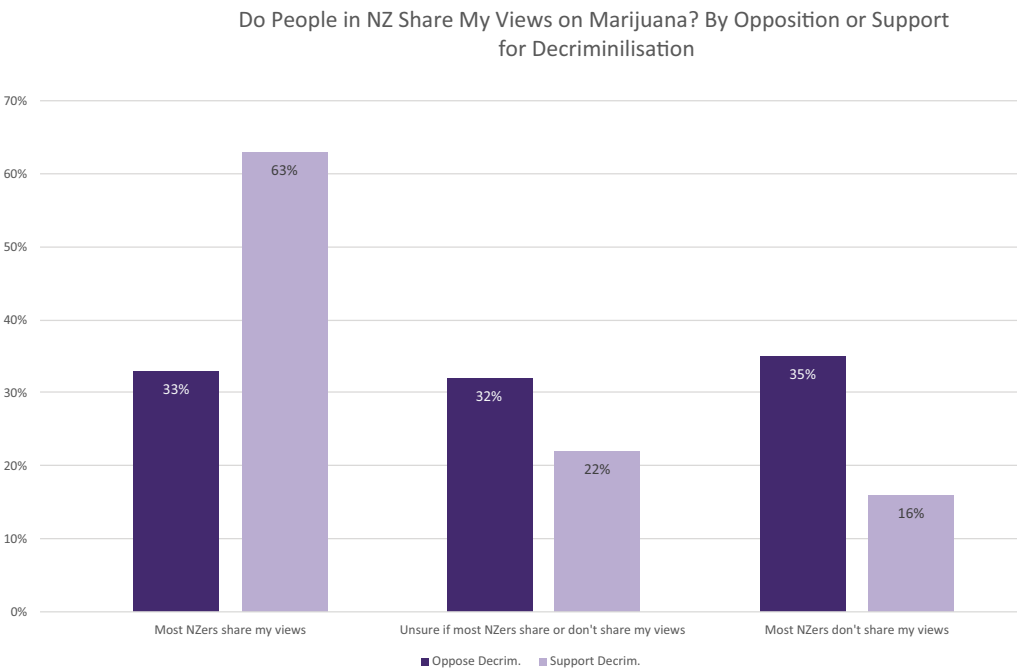
We explored if either of these phenomena occurred in New Zealand with regard to our topic of the decriminalisation of marijuana.

In New Zealand, in this survey, those who oppose or strongly oppose decriminalisation represent the minority (35%) while the majority (58%) strongly or somewhat support its decriminalisation.

Just over a third of those holding the minority views - 35% - correctly identified that most other people in New Zealand didn't share their views. Another third incorrectly identified and another third were unsure. This finding suggests there is a range of beliefs about how widely held the minority opinion is on this topic in New Zealand. There is no particularly clear false consensus effect

Conversely, of those who supported decriminalisation, most (63%) identified their views were shared by other New Zealanders. Showing a fairly accurate assessment of public opinion.

The findings suggests that New Zealanders do, despite the echo chamber, polarising and amplification effects of digital media, gain relatively accurate views of others opinions on the decriminalisation of marijuana. Participants stated they receive most information on the topic via mainstream media sources and at the time of this survey there had been several publicly published and widely reported opinion polls. Together these factors, along with our small population size, may factor in people’s ability to accurately identify they held a minority view on the issue.



PEOPLE WHO ARE UNWILLING TO DISCUSS A POLITICAL ISSUE (DECRIMINALISATION OF MARIJUANA) IN REAL LIFE SETTINGS ARE ALSO UNWILLING TO DO SO ONLINE.

The results of this survey suggest that the existence of online forums does not increase people’s willingness to participate in democracy. People who were unwilling to discuss the decriminalisation of marijuana at a community meeting, workplace, family dinner were also unwilling to do so on Facebook. While those who were willing to discuss the issue in these settings were also willing to do so online.

For example, we found that those willing to join in a conversation about decriminalising marijuana at a community meeting were significantly more likely to be willing to join in a conversation about decriminalising marijuana on Facebook (67%). Those unwilling to join in a conversation about decriminalising marijuana at a community meeting were significantly less likely to be willing to join in a conversation about decriminalising marijuana on Facebook (20%).

These findings replicate findings from the Pew Research centre in the US.¹

THERE IS EVIDENCE THAT NEW ZEALANDERS WHO BELIEVE THEIR VIEWS ARE IN THE MINORITY ON DECRIMINALISATION OF MARIJUANA ARE LESS WILLING TO SHARE THEIR VIEWS BOTH OFFLINE AND ONLINE.

A spiral of silence refers to the idea that people are unwilling to speak up on issues when they believe most people in society don't hold similar views. It was originally proposed by Elisabeth Noelle-Neumann in 1974. We explored whether people's perceptions about the majority view on the legalisation of marijuana affected their willingness to discuss it both offline or online.

We asked people whether they thought their friends and family shared their views on decriminalisation and whether the rest of NZ held similar views. We found a clear relationship between people's willingness to discuss the issue in offline settings and their perceptions of their friends and families views.

For example, we found that those who believe their friends and family held different views showed significantly greater hesitancy to join in a conversation about decriminalising marijuana at work (32%). Those who believed their friends and family shared their views were significantly less likely to show a hesitancy about joining in a conversation about decriminalising marijuana at work (15%). In other words a belief that views were shared increased willingness to share them at work, while a belief they were not lead to unwillingness to share. The same pattern was found for all offline settings.

The pattern was different for online discussions. Those who believed that family or friends shared their views were significantly more willing to discuss the issue on Facebook, in other words a belief in shared views offline lead to a willingness to discuss it online. However, those who believed friends and family held different views were still willing to discuss the issue on Facebook. This finding suggests that social media may, for some people, be a tool to share their views when they believe family and friends don't agree with them. We did not ask where on Facebook these views were shared, it may be that Facebook groups offer a space to discuss views not shared within a person's direct social group.

What about perceptions of the rest of New Zealand? Did this impact willingness to discuss the issue offline or online? People who believe other New Zealanders held different views were significantly more hesitant to discuss the issue at a community meeting and at a restaurant with friends, but not at work, or a family dinner. In other words, a belief that others held different views did not reduce willingness to discuss that issue in some in person settings.

People who thought their views were different from the rest of New Zealand were significantly more hesitant to discuss the issue on Facebook (58%). Those who thought their views were shared were less hesitant to discuss them on Facebook (37%) - in other words a belief that views were shared by the majority of people lowered barriers to discuss the issue online.

See table 1 below

If the topic of decriminalising marijuana for personal use came up in the following places, how willing would you be to join in the conversation? — On Facebook

		Very willing	Somewhat willing	Total willing	Somewhat unwilling	Very unwilling	Total unwilling	Unsure	Row population
Total agree	%	19%	43%	61%	23%	14%	37%	2%	431
Total disagree	%	7%	32%	39%	27%	31%	58%	3%	195
Unsure	%	5%	18%	23%	29%	36%	65%	12%	240
NET	%	12%	33%	46%	25%	24%	49%	5%	867

CONCLUSIONS

In this short survey we explored some attitudes and behaviours of a representative group of New Zealanders to social media, in the context of issues of democracy.

The large majority of people surveyed used social media, with Facebook being the platform most frequently named. There is still a small group of the population, who are more likely to be older men, not using any social media, which is an important finding.

Around a quarter of people who used social media used it for explicitly named “political and social issues”. Because we did not prompt people or define the term, it is possible that people very narrowly defined the term political or social issue. For example as something explicitly identified as involved with politics, so care should be taken with this finding. We did find differences in who used social media to engage in democratic activities, specifically greater proportions of people from minority ethnic groups use social media to engage with democratic issues in New Zealand than NZ European/Pākehā, which is an important finding in terms of the access that social media gives to those traditionally excluded from political engagement.

We found a low trust environment of news shared online. People were clearly aware of problems with accuracy and misinformation from news shared online - indicated by the majority of people agreeing that they always checked the accuracy of news shared online regardless of the source. However, people also showed high trust in their friends and family as sources of information. Given the biases inherent in search engine algorithms and other online media platforms (for example YouTube), and that misinformation and disinformation spreads faster and further than accurate information online, using trusted social connections is a problematic technique to counter inaccuracy in online news.

People expressed much higher trust in print and mainstream media sources (TV and Radio). This presents an opportunity to ensure that these news sources are sufficiently funded to convey the best, most accurate information, in ways that take account of the power and spread of misinformation across social media. There is a clear case to be made for these sources to employ evidence-based techniques to actively overcome disinformation and misinformation.

We explored peoples opinions on the decriminalisation of marijuana and the role digital media may have in informing them. We chose this topic not because we were interested in the particular issue, but because it represented a public policy issue that had high levels of media exposure, that was moderately polarising, but was not generally viewed as highly polarising. We found most people supported its decriminalisation, and receive information about it from mainstream news sources (not digital media). There was no evidence that large numbers of people were inaccurately informed about others’ views on the issue. Though around a third of those who held a minority view on the issue incorrectly believed most other New Zealanders shared their views.

In this survey, social media does not provide an alternative setting to discuss a political issue for those people who are unwilling to discuss the decriminalisation of marijuana offline. In this sense social media looks like it replicates other offline social settings.

We found evidence of spirals of silence in relation to the decriminalisation of marijuana, most especially in relation to people’s perceptions of what other New Zealanders thoughts (as opposed to their friends and family). Our finding replicates international findings. Social media in New Zealand, notably Facebook may not be a forum in which people who believe they hold minority views feel they can discuss those views more easily.

APPENDIX 1

METHODOLOGY

The report is based on results from an online survey among a nationally representative sample of New Zealanders aged 18 years and over.

The sample size was n=1,000 and fieldwork was carried out from the 27th of September to 2nd of October 2018.

The survey data was rim weighted by region, age, gender and ethnicity. Rim weighting is designed to weight characteristics simultaneously. The rim weighting process attains all the desired proportions while distorting each variable as little as possible.

The margin of error for a 50% figure at the 95% confidence level for a sample size of 1,000 is plus or minus 3.1%.

SURVEY QUESTIONS

Please tell me if you ever use...

- ☐ Twitter
- ☐ Facebook
- ☐ Other social media platform
- ☐ Neither [SINGLE]
- ☐ Unsure [SINGLE]

Do you currently...

- ☐ Follow any elected officials, candidates for office or other political figures on a social media platform.
- ☐ Belong to a group on a social media platform that is involved in political or social issues, or that is working to advance a cause (e.g. climate change, conservation, gender or health issues)
- ☐ Neither [SINGLE]
- ☐ Unsure [SINGLE]

Have you ever used social networking sites to...

- ☐ Post links to political stories or articles for others to read
- ☐ Post your own thoughts or comments on political or social issues
- ☐ Sign or share online petitions on a political or social issue
- ☐ Encourage other people to take action on a political or social issue
- ☐ Engage with politicians online
- ☐ Encourage other people to vote
- ☐ Repost or share content related to political or social issues that was originally posted by someone else
- ☐ "Like" material related to political or social issues that others have posted
- ☐ None of the above [SINGLE]
- ☐ Unsure [SINGLE]

How much do you agree with the following ...

Strongly agree | Agree | Disagree | Strongly disagree | Unsure

- ☐ I trust news stories that my friends share online
- ☐ I distrust most news that people share online
- ☐ I assess the accuracy of online news depending on how much I trust the person who shared it
- ☐ I use other sources to check the accuracy of online news myself, no matter who shares it with me

[ASK ALL]

How much do you trust information from

Trust lot | Somewhat trust | Somewhat distrust | Completely distrust | Unsure

- ☐ Your local print newspaper
- ☐ TV and radio
- ☐ Friends and family
- ☐ [If 1b = yes ASK] Facebook
- ☐ [If 1a = yes ASK] Twitter
- ☐ Online news sources excluding Facebook/Twitter

The next few questions are about the potential of decriminalising the personal use of marijuana in New Zealand (this will mean New Zealanders will be able to grow marijuana for personal use without being prosecuted).

Recently, decriminalising of marijuana for personal use has been in the news.

How interested are you, if at all, in this topic?

- ☐ Very interested
- ☐ Somewhat interested
- ☐ Not that interested
- ☐ Not interested at all
- ☐ Unsure

Overall, how knowledgeable would you say you are about the debate surrounding the decriminalising of marijuana for personal use?

- ☐ Very knowledgeable
- ☐ Somewhat knowledgeable
- ☐ Not that knowledgeable
- ☐ Not knowledgeable at all
- ☐ Unsure

How much have you read, seen or heard about the debate surrounding decriminalising of marijuana for personal use from....

A lot | Some | A little | No information | Unsure

- ☐ Your local print newspaper
- ☐ TV and radio
- ☐ Friends and family
- ☐ [If 1b = yes ASK]: Facebook
- ☐ [If 1a = yes ASK]: Twitter
- ☐ Online news sources excluding Facebook/Twitter

[If any of 7d, e or f = 'a lot', 'some' or 'a little' ASK]

From which of the following sources did you get information about the debate surrounding the decriminalising of marijuana for personal use?

- ☐ Herald.co.nz
- ☐ Stuff.co.nz
- ☐ Newsroom
- ☐ The Spinoff
- ☐ The Wireless
- ☐ Whaleoil
- ☐ The Daily Blog
- ☐ Other please write:
- ☐ [SINGLE] Unsure [SINGLE]

Thinking about the debate over the decriminalising of marijuana for personal use ... Do you FAVOUR or OPPOSE this happening in New Zealand?

- ☐ Strongly favour
- ☐ Somewhat favour
- ☐ Somewhat oppose
- ☐ Strongly oppose
- ☐ Unsure

If the topic of decriminalising of marijuana for personal use came up in the following places, how willing would you be to join in the conversation?

Very willing | Somewhat willing | Somewhat unwilling | Very unwilling | Unsure

- ☐ At a community meeting
- ☐ At work
- ☐ At a restaurant with friends
- ☐ At a family dinner
- ☐ [If 1b = yes ASK] Facebook
- ☐ [If 1a = yes ASK] On Twitter

Still thinking about the current debate about decriminalising marijuana for personal use, to what extent do you agree with each of the following ...

Strongly agree | Agree | Disagree | Strongly disagree | Unsure

- ☐ Most of my friends and family hold similar views as me about decriminalising marijuana
- ☐ Most of New Zealand hold similar views as me about decriminalising marijuana

Still thinking about the current debate about decriminalising marijuana for personal use, how polarising do you think this issue is in New Zealand?

- ☐ Very polarising
- ☐ Somewhat polarising
- ☐ Not that polarising
- ☐ Not polarising at all
- ☐ Unsure

Which gender do you identify as:

- ☐ Male
- ☐ Female
- ☐ Different identity (please specify)
- ☐ Prefer not to say

Which of these age groups do you fall into?

- ☐ 17 years and under
- ☐ 18-19
- ☐ 20-24
- ☐ 25-29
- ☐ 30-34
- ☐ 35-39
- ☐ 40-44
- ☐ 45-49
- ☐ 50-54
- ☐ 55-59
- ☐ 60-64
- ☐ 65-69
- ☐ 70-74
- ☐ 75 and over
- ☐ Prefer not to say

Which of the following regions best describes where you live?

- ☐ Northland Region
- ☐ Auckland Region
- ☐ Waikato Region
- ☐ Bay of Plenty Region
- ☐ Gisborne Region
- ☐ Taranaki Region
- ☐ Manawatu-Whanganui Region
- ☐ Hawke's Bay Region
- ☐ Wellington Region
- ☐ Marlborough Region
- ☐ Tasman/Nelson Region
- ☐ West Coast Region
- ☐ Canterbury Region
- ☐ Otago Region
- ☐ Southland Region

Which of the following best describes which area of Auckland you live in?

- ☐ **Central** (Balmoral, Epsom, Greenlane, Kingsland, Maungawhau, Mt Eden, Morningside, Mt Albert, Mt Eden, One Tree Hill, Owairaka, Pt Chevalier, Sandringham, Waterview, Hillsbrough, Lynfield, Mt Roskill, Three Kings, Waikowhai, Wesley, Ellerslie, Glendowie, Kohimarama, Meadowbank, Mission Bay, Orakei, Remuera, St Heliers, St John, Glen Innes, Mt Wellington, Onewa, Oranga, Panmure, Penrose, Pt England, Royal Oak, Te Papapa, Auckland CBD, Grey Lynn, Herne Bay, Newmarket, Parnell, Ponsonby, St Mary's Bay, Westmere)
- ☐ **Hauraki Gulf Islands** (Waiheke, Great Barrier etc.)
- ☐ **West** (Glendene, Henderson, Massey, Ranui, Te Atatu, West Harbour, Westgate, French Bay, Glen Eden, Laingholm, O'Neill Bay, Oratia, Titirangi, Waikaraka, Waitakere, Whatipu, Wood Bay, Woodlands Park, Avondale, Blockhouse Bay, Green Bay, Kelston, New Lynn, New Windsor, Rosebank, Waterview)
- ☐ **North** (Helensville, Huapai, Kawau Island, Kumeu, Matakana, Warkworth, Wellsford, Browns Bay, Campbells Bay, Mairangi Bay, Orewa, Silverdale, Waiwera, Whangaparaoa, Albany, Greenhithe, Hobsonville, Northcross, Paremora, Pinehill, Wainoni, Whenuapai, Castor Bay, Devonport, Stanley Bay, Sunnynook, Takapuna, Beach Haven, Birkdale, Birkenhead, Chatswood, Glenfield, Hillcrest, Marlborough, Northcote)
- ☐ **South** (Franklin, Favona, Mangere, Otahuhu, East Tamaki, Manukau Central, Middlemore, Otara, Papatoetoe, Puhunui, Homai, Manurewa, Wattle Downs, Weymouth, Wiri, Alfriston, Drury, Hingaia, Pahurehure, Papakura, Red Hill, Takanini, Ardmore, Awhitu, Beachlands, Clevedon, Karaka, Kawakawa Bay, Maraetai, Orere Point, Pukekohe, Waiuku, Whitford)
- ☐ **Howick Ward** (Botany, Bucklands Beach, Cockle Bay, Dannemora, East Tamaki, Eastern Beach, Farm Cove, Flat Bush, Half Moon Bay, Highland Park, Howick, Mellons Bay, Northpark, Pakuranga, Shelly Park, Somerville, Sunnyside)

Generally speaking, do you usually think of yourself as a supporter of:

- ☐ National
- ☐ Labour
- ☐ Green
- ☐ NZ First
- ☐ ACT
- ☐ Māori party
- ☐ Other party
- ☐ No party
- ☐ Prefer not to say

Which of the following ethnic groups do you belong to?

One or several groups may apply to you.

- ☐ NZ Māori
- ☐ NZ European
- ☐ British
- ☐ Other European
- ☐ Pacific Island
- ☐ Chinese
- ☐ Indian
- ☐ Other Asian
- ☐ Other (please specify):

the 1990s, the number of people in the world who are under 15 years of age is expected to increase by 1.5 billion, from 1.1 billion in 1990 to 2.6 billion in 2010.

As the world's population grows, the demand for food and other resources will increase. The world's population is expected to reach 9 billion by 2050, and the demand for food will increase by 50% by 2030.

The world's population is also becoming more urban. In 1990, 54% of the world's population lived in urban areas. By 2010, this is expected to increase to 69%.

The world's population is also becoming more diverse. In 1990, 54% of the world's population was of European descent. By 2010, this is expected to decrease to 44%.

The world's population is also becoming more educated. In 1990, 54% of the world's population was illiterate. By 2010, this is expected to decrease to 44%.

The world's population is also becoming more mobile. In 1990, 54% of the world's population lived in rural areas. By 2010, this is expected to decrease to 44%.

The world's population is also becoming more affluent. In 1990, 54% of the world's population lived on less than \$2 a day. By 2010, this is expected to decrease to 44%.

The world's population is also becoming more healthy. In 1990, 54% of the world's population lived in poor health. By 2010, this is expected to decrease to 44%.

The world's population is also becoming more peaceful. In 1990, 54% of the world's population lived in a state of conflict. By 2010, this is expected to decrease to 44%.

The world's population is also becoming more democratic. In 1990, 54% of the world's population lived in a non-democratic country. By 2010, this is expected to decrease to 44%.

The world's population is also becoming more environmentally conscious. In 1990, 54% of the world's population lived in a country that was not a member of the United Nations Environment Programme. By 2010, this is expected to decrease to 44%.

The world's population is also becoming more technologically advanced. In 1990, 54% of the world's population lived in a country that did not have a telephone. By 2010, this is expected to decrease to 44%.

The world's population is also becoming more globally connected. In 1990, 54% of the world's population lived in a country that was not a member of the World Trade Organization. By 2010, this is expected to decrease to 44%.

The world's population is also becoming more environmentally sustainable. In 1990, 54% of the world's population lived in a country that was not a member of the United Nations Sustainable Development Goals. By 2010, this is expected to decrease to 44%.

The world's population is also becoming more socially responsible. In 1990, 54% of the world's population lived in a country that was not a member of the United Nations Human Rights Council. By 2010, this is expected to decrease to 44%.

The world's population is also becoming more environmentally friendly. In 1990, 54% of the world's population lived in a country that was not a member of the United Nations Framework Convention on Climate Change. By 2010, this is expected to decrease to 44%.

The world's population is also becoming more environmentally aware. In 1990, 54% of the world's population lived in a country that was not a member of the United Nations Convention on Biological Diversity. By 2010, this is expected to decrease to 44%.

APPENDIX 4: REPORT ON QUALITATIVE INTERVIEWS

D I G I T A L T H R E A T S T O D E M O C R A C Y



> > R E P O R T > > > >

> > > > O N > > > > > >

> Q U A L I T A T I V E >

> > I N T E R V I E W S >



Luminate

D I G I T A L T H R E A T S T O D E M O C R A C Y

This report is part of the Digital Threats to
Democracy research project.

To see the rest of the reports and the
overall findings go to digitaldemocracy.nz

ISBN: 978-0-473-48026-4

CONTENTS

BACKGROUND, PURPOSE AND METHOD	147
Research purpose and questions	147
Methodology	148
Definitions	148
WHAT IS THE LANDSCAPE OF DIGITAL DEMOCRACY?	150
What hasn't changed?	150
What has changed?	153
What's different about NZ, compared to other countries?	156
WHAT ARE THE OPPORTUNITIES?	158
Democratisation of information	158
Improving participation in democracy	159
Transparency in government	162
A messy picture	165
WHAT ARE THE RISKS AND PROBLEMS?	166
Digital exclusion	166
Monopolies and lack of competition	168
Impact on public-interest media	170
Misinformation, Disinformation and Malinformation	171
Political manipulation including foreign interference	176
Online abuse	181
Silencing	184
Disengagement and attention hijacking	185
Loss of privacy	189
Polarisation, radicalisation and 'echo chambers'	191
Lack of transparency and autonomy	193
SOLUTIONS	196
Change structures & systems	196
Change context & environment	206
Long lasting protections	217
Change understanding & behaviours	218
WHAT ROLE FOR NZ?	220
Follower	221
Leader	221
Niche influencer	222

BACKGROUND, PURPOSE AND METHOD

RESEARCH PURPOSE AND QUESTIONS

Digital media has been heralded as inherently democratising. People have direct access to each other and to their elected representatives, across geographical and cultural boundaries. But increasingly it is also seen as a space in which democracy may be simultaneously undermined.

As digital tech increasingly permeates society, there is good reason to pay attention to the institutions, policies, and practices that surround this technology which present both opportunities and threats to democracy. This is especially true for government, but is also true for everyone with an interest in the future health of our democracy.

The purpose of this research was to explore the opportunities, risks and threats posed to New Zealand's democracy by digital media, in order to scope out future research into the policy solutions available to New Zealand to maximise the opportunities, and to meet and mitigate the threats.

The research questions for this project were:

1. What are the opportunities, risks and threats to democracy in New Zealand posed by digital development, and in particular the pervasiveness of social media and the digital platform monopolies?
2. How are these technologies impacting on democracy? Through what processes/platforms etc, and how does that have an impact on democracy? For whom?
3. What opportunities, risks and threats do these technologies present specifically for building a more equitable and inclusive society?
4. Who are the key experts and stakeholders (including those historically impacted by inequity?) with an interest in this issue?
5. What other work is currently being done in this area to ensure that New Zealand is equipped to respond to those threats?
6. What, therefore, are the most relevant, urgent and useful aspects of this issue to be explored further?

METHODOLOGY

A series of in-depth interviews were conducted with experts and stakeholders to explore the scope of this issue in more detail, prioritise various aspects of the problem for future research and identify key potential collaborators for further research.

Thirty five in-depth interviews were undertaken with a selection of experts, stakeholders and users drawn from the following sectors.

- > Political
- > Policy and official (including Department of Internal Affairs, Human Rights Commission)
- > Iwi and Māori-led organisations
- > Civil society (including InternetNZ)
- > Industry/sector organisations (including digital platform, data analytics, media, data security, digital marketing)
- > Academics, researchers and experts (including legal, information security and privacy, media, economics, tikanga Māori)
- > International experts.

The interviews were recorded and transcribed and analysed using a hybrid of content and grounded analysis in which some broad themes were used as a starting framework for the analysis, but amended and altered based on the themes that emerged from the data as the analysis progressed. Participants were given the opportunity to see the quotes attributed to them. Some participants chose not to have their quotes attributed to them, but most people agreed to be named.

DEFINITIONS

In order to assess the impact of something like digital media on democracy, you need some sort of definition of democracy. Which, it turns out, is easier said than done. From the outset we were clear that we needed a broad definition of democracy, which included not only electoral democracy and voter participation in elections, but also the ability of citizens and residents to access accurate information, and participate in public discussions about policy issues and politics. Freedom of speech, freedom of political expression and freedom from discrimination in public life are all important components of democracy, as is the ability of citizens to be informed, participate in discussions and influence decisions about public issues that affect them.

In the end, we used a definition of democracy adapted from the framework developed by the Economist Intelligence Unit for their Democratic Index report, and the definition used by Jamie Bartlett in his book 'The People vs Tech'.

The five features of democracy in our definition are:

- > **Electoral process and pluralism:** including whether elections are free, fair and trusted.
- > **Active citizens:** alert, informed citizens who are capable of making important moral judgements. This includes measures of equity and diversity in representation.
- > **Shared democratic culture:** whether there's enough societal consensus, cohesion and willingness to compromise for a stable, functioning democracy. In New Zealand, this would include compliance with te Tiriti o Waitangi, on which our shared democratic culture is founded.
- > **Civil liberties and competitive economy:** a functioning competitive economy and civil society, including protection of human rights and free, independent media.
- > **Trust in authority:** government, parliament and judiciary are trustworthy and elected representatives are accountable to the people.

Defining democracy is complex. Defining digital media is almost as difficult. Digital media technically includes all digitised content that can be transmitted over the internet or computer networks. This could include text, audio, video, and images. So content from print or broadcast media outlets can fall into this category when it is presented on a website or blog.

The focus of this research was on social media, online forms of communication that people use to share and exchange information with interested audiences, and within that, a specific focus on the major digital platforms, including Facebook, YouTube and Twitter. However, interviewees also talked about the impact of other forms of digital media on democracy, including blogs, online forums and digital forms of traditional media.

WHAT IS THE LANDSCAPE OF DIGITAL DEMOCRACY?

One of the risks in assessing the impact of digital media on democracy is that we overlook the many other factors affecting the health of our democracy, and attribute too much influence to the role of digital media. So we start our analysis with a consideration of what has not changed: what are the challenges and opportunities for democracy in New Zealand which predate and are not dependent on the rise of digital media?

WHAT HASN'T CHANGED?

MISINFORMATION, DISINFORMATION AND HARASSMENT ARE NOT NEW.

When asked whether the growing ubiquity of digital media as a means by which New Zealanders are accessing information was undermining the quality of the information that people are using to form their opinions on matters of public interest, some interviewees said that, by and large, it was not.

Media commentator and journalist, Colin Peacock didn't think digital media was having a huge impact on the quality of information available to New Zealanders, "although just this week we're looking at this issue of 1080, where it's clear that some thousands, possibly tens of thousands, of people are reading a lot of content which is excluding important scientific information." As Peacock's colleague Jeremy Rose pointed out, misinformation existed well before the internet. "I think if you'd done a survey ten years ago, before the internet was as big as it is now, that you would actually have a higher percentage [of people who were misinformed]. So I'm skeptical. They're definitely using it [to spread misinformation], but so is the other side, and that is a democratising kind of thing."

Political activist, blogger, and pollster David Farrar pointed out that the spread of misinformation 'used to happen in the pub' and continues to be spread offline. "Let's take the rumours about Clarke Gayford. Even though the internet played a small part in spreading them, that spread up and down NZ through word of mouth. I had people coming up to me in the street. It just seemed to be everywhere." Although he added that

the internet has amplified the spread of that kind of misinformation, and has made it very easy to do.

Former politician Marama Fox also noted that abuse and hate can be, and always has been, communicated without any recourse to the internet or digital media. “I got threatened the old fashioned way too, by straight up, anonymous, handwritten mail.”

OUTRAGE, POLARISATION AND EXTREMISM ARE NOT NEW

Other interviewees remarked that an appetite for information that is shocking or outrageous also predated the rise of digital media. Dave Moskovitz, a member of the Ministerial Advisory Group on Digital Economy and Digital Inclusion explains: “Media has always been about the shocking, the unusual, the thrilling. Stories of people selflessly doing things for other people are just not interesting. You’ll never see them in the news.”

Digital communications consultant Gwynn Compton agrees that the human appetite for shocking content predates the internet and also argues that the use of digital media by extremist groups is not so much a new phenomenon as the newest incarnation of an historical pattern.

“When you look through political history of far right movements, it’s always been the thing that they’ve exploited: the new, emerging communication channels to push whatever message that they want to. In the 20’s and 30’s it was radio, it was the big thing that they pushed really effectively. And now it’s social media, and search advertising seems to be the new barrel. And all the messages are fundamentally the same. We know it’s dangerous from what we’ve seen in the 1920’s-30’s Europe, and you can see similar stuff happening now with the rise of far-right parties.”

On the other hand, Compton points out that digital media has made it easier for people with harmful or extreme views to connect with each other, “now all the racist uncles in the country can find each other online and start to organise. That emboldens them, they don’t feel they’re such a small number, even though across the country they still are.”

THE GUTENBERG PARENTHESIS AND THE SHORT REIGN OF THE AUTHORITATIVE TEXT

The Gutenberg Parenthesis is an expression that refers to the idea that from the start of publishing to the start of internet publishing was a short period of time during which there was a shared public belief in the authority of published material. This period falls between the invention of the press (by Johannes Gutenberg) and the invention of the internet. These two events form a parenthesis around what is understood, in this idea, to have been an exceptional time during which people assumed a level of authority in all published material. This authority came from the fact that someone had curated material, assessed it, said it was worth saying, and someone had bought it and published it which had validated that assessment.

As InternetNZ’s Jordan Carter explained, before publishing, we had storytelling, and

after internet publishing, we have a situation where anyone can publish anything. “So it’s the story of the rise and fall of authoritativeness in written material. I think the mass broadcast media played, for a short time, a unifying role, and maybe a democratising role. You know, people sitting around the TV, watching the news. So, it seems to me that we might be going back to something that we were a bit more like, anyway? A bit more of the diversified set of perspectives.”

FILTER BUBBLES, SOUNDBITE POLITICS AND DATA CAPTURE - NONE OF THIS IS NEW

In a similar vein, Carter explained that during his time in London he noticed something that could be seen as a pre-digital version of the filter bubble: “You could judge people based on the title of the newspaper they were carrying under their arm. ‘Oh, that’s a Sun person, I don’t want to talk to them.’ I think that desire for affiliation, is just a part of who we are and the role of the mass broadcast media in providing a common unifying perspective, was a relatively short term thing.”

Member of Parliament Chloe Swarbrick reflected that politics today operate on the ‘soundbite’, a pattern sometimes attributed to the influence of digital media, but she also questioned whether this was a new phenomenon. “I don’t know if we ever didn’t [work in soundbites]. I don’t know if people used to hang out to watch full debates of stuff, or if they used to engage in full, you know, reading long articles which explain both sides in detail.”

Elyse Rob  rt, a digital media researcher and practitioner, also pointed out that humans have always gathered data. “I’ve been doing some genealogy research and I can go back to 1800s Australia and see that a great-great-grandparent’s husband abandoned them. His age, weight and occupation is all in there. We have been collecting and sharing gossip and data through media for centuries.” Although, she concedes, data collection is “perhaps quicker and easier to disseminate” in the digital age.

Rob  rt also points out that human agency hasn’t changed in the digital age, “just because people are being served more of one flavor than the other, that doesn’t mean that they have lost their agency, their literacy and their power to critique.”

Finally Rob  rt argued that the harmful impact of digital media on social cohesion has been exaggerated. “Our communities have emerged online, extended into the world wide web, which is still real. It is a different mode of interaction, with different social codes, but whether we have actually lost our sense of community or simply changed its shape is variable from space to space.”

CULTURAL HEGEMONY OF TECH ISN’T NEW

Erika Pearson is a Senior Lecturer in Communication at Massey University with a particular interest in digital culture, social media and online social networks. She argues that the cultural and ideological perspectives of the people who dominate the tech sector has influenced the design of digital media.

“The hetero-masculine, North American, white, middle class ideology of Silicon Valley is replicated through technology. The Silicon Valley generalists are engineers from Ivy

League schools who have a fairly toxic masculinity in a very particular historical, economic and social context, who feel that their view is the norm and everything else is deviation.”

The argument is that when these engineers build software, they build it to meet the needs of themselves and people like them. The design testing of the early platforms, says Pearson, involved calling up “three of their mates, have some beers, and test the software – oh look, it works for us, therefore it must work for everyone.” Then that software would be deployed to a general population which included people with very different perspectives, experiences and needs. Including, says Pearson, “trauma survivors, veiled women, indigenous communities”, who then have to negotiate how they use a platform that wasn’t designed with their needs in mind.

One of the barriers to change identified by Pearson is that the digital media companies have a “very paternalistic” model of management, in which the companies themselves, run largely by the same narrow group of people who built the software, have to determine harm. “For these companies, you can’t tell them you have experienced harm, they have to tell you whether or not you have experienced harm. And they are telling you no, that experience was just a misunderstanding or just a bit of robust conversation you shouldn’t take personally.”

Pearson’s arguments are reflected in the comments of other interviewees who observe the irony in teams of predominantly white, middle-class men trying to create something ‘disruptive’, who nevertheless managed to build platforms that replicates the social structures of the 1950s. It could be argued that despite all that has changed as a result of digital media, this replication of existing power structures in the governance and management of the tech giants inevitably lead to reinforcing many already entrenched power imbalances. Further, that the lack of diversity at the governance and senior management level prevented these companies from identifying and responded adequately to the risks and threats inherent in their platform designs.

WHAT HAS CHANGED?

While recognising that the foundations of mis- and disinformation, online harassment and abuse, polarisation and extremism all existed well before the rise of digital media, most interviewees nonetheless saw particular ways in which the features and functions of digital media has changed the scale, intensity and reach of those phenomena.

SCALE, SPEED, BREADTH AND INTENSITY

One of the most obvious ways in which digital media differs substantially from previous forms of media and communications is its speed and breadth of reach. As Pia Andrews, digital and open government pioneer and former Service Integration Lead at the Department of Internal Affairs explains, “It sped it up. It dramatically sped it up. It’s the speed and it’s the breadth. Someone can share something silly, not just with their friends now, but with the whole world.”

This speed and breadth of reach are critical factors in the way, for example, misinformation and disinformation can be spread digitally. It was also horribly illustrated in the speed with which the livestream video of the Christchurch mosque massacres made its way to viewers around the world, and in how difficult it has been for authorities in New Zealand to remove the video from the internet and pursue charges against the people

who had shared it, even with the cooperation of the digital platforms.

In the context of harassment and abuse the most common view expressed by experts was that digital spaces amplified and intensified abuse. David Farrar observed that digital media provides the means for people to be publicly shamed at a speed and scale that is unprecedented, “I think of the woman who was flying to Africa and said, “I won’t get AIDS, I’m white.” It was an incredibly crass, stupid thing to say, her defense was it was meant to be sarcasm. But regardless, by the time she’d landed there were 50 people at the airport with cameras to record her turning her phone on. She became a global pariah.”

A related change in the intensity of abuse and harassment highlighted by Sandra Dickson, Program Lead at Gender Equal NZ when this research was carried out, is that digital media enabled people to engage with others without having to actually witness the impact of their abuse, or look their targets in the eyes. “We’re not having to respond to the persons face when we write the disgusting comment about them.” She also reflected that digital spaces can lack the kind of dissenting voices that might temper what she described as “the pile-on culture that you get in the digital world”.

MICROTARGETING

Another factor that differentiates digital media from what came before it is the ability it gives to target advertising and information to particular audiences. As technology reporter Ian Apperley noted, digital advertising on platforms like Google and Facebook can be targeted using the personal data gathered by those platforms. He gave an example in which he had been talking with someone about traveling to Japan.

“The moment we start having those conversations, I start getting Tokyo Air NZ specials in my media feeds. In the NZ Herald, in my Facebook feed, and so forth. So Google’s moving a little ahead of the game in terms of [gathering data from] the conversations that you and I are having, and targeting advertising on that basis.”

Apperley argued that this ability to personalise advertisements to users combined with the ability to make targeted advertising look like user content, both of which are features of digital marketing, made it easier to influence people. “And frankly, it’s incredibly easy to do. Facebook’s perfect for that, because it allows for targeted advertising. So I can pick all white males over the age of 50, who have a slightly right-leaning and potentially religious background, and then I can push right-wing stuff to them, knowing that they will push that out without even thinking, to their family, and friends, and so forth. It’s terrifying.”

While he acknowledged that advertising has always been in the business of “messing with people’s minds,” Apperley also argued that the power of traditional advertising was dwarfed by the potential of advertising in the digital age, “a static image, or an ad that’s inserted into a break on a TV show is nothing compared to what we have now.

AN AGE OF DISTRACTION. AN AGE OF OUTRAGE.

Interviewees also identified a range of other changes to our democratic environment which could be attributed to the rise of digital media. In particular, they pointed to

the ‘attention overload’ generated by social media as a factor influencing our ability to concentrate on political issues and communicate productively about them. Author Nicky Hager identifies the “total and constant disruption” of social media as one of its primary impacts on democracy. The kind of deep thinking and reflection needed for an informed citizenry to engage on complex social, environmental and political issues is, says Hager, almost impossible in the face of this constant distraction.

Another theme that comes up when you talk about the impact of digital media on democracy is the role of outrage. Commentators have dubbed this the ‘age of outrage’, citing the role of social media in generating and spreading outrage. Interviewees said that the design of digital platforms plays a role in determining the nature of the debates and discussions that take place on those platforms.

Other interviewees focused on the underlying business model of the major platforms, which reward people for engagement, no matter the nature of that engagement. Because their profitability depends on the ability to attract and retain people’s attention, digital platform companies employ teams of psychologists who design, test and iterate features that provide almost irresistible physiological and emotional rewards for getting people to engage with your content.

DATA AS THE NEW OIL

In relation to privacy, interviewees noted that users’ personal data is increasingly being mined to sell to advertisers, often without conscious consent from users who are worn down by ‘consent fatigue.’ Digital platforms, and the myriad applications created to integrate with them, are designed to give users incentives to trade in their personal privacy for the efficiency and gratification provided by these online spaces.

Data has been referred to as the new oil, a phrase credited to Mathematician Clive Humby in 2006, and given new currency since the Economist published a 2017 report asserting “the world’s most valuable resource is no longer oil, but data.” Some have argued with this, but overall there is agreement amongst experts that the extraordinary growth, and profitability, of the platform giants has been built on their accumulation of unprecedented amounts of personal data about their users. Those users, however, have not always been aware that their data was being collected and sold.

Economist Ganesh Nana explains that a central assumption in economics is that a market exchange is made from an informed position, consumers have information and make decisions based on that information and so do producers. “The theory of the market reaching an optimal position benefiting everybody is based on the assumption that consumers make their decisions from an informed basis; they know the value of the thing that they want and they know the consequences of the decisions they are about to make and there are no externalities.” But in the case of digital media platforms, he says, most consumers don’t have good information about the product they are using and the personal data they are trading in exchange for that use.

“If we don’t have good information,’ says Nana, “we will make bad decisions and reach what we call a sub-optimal position ... So, there is a market failure. Because how many how many consumers actually understand that they are paying for it with their data? I would say a relatively small percentage of consumers that actually make that connection.”

DOMINANCE OF A HANDFUL OF COMPANIES

This market failure described by Ganesh Nana, was reflected by interviewees who focused on the extraordinary dominance in the market of a handful of digital platform companies, including Facebook and Google. The power and market dominance of these companies surpasses that of the mainstream media monopolies of previous decades.

Dr Marcin Betkier is a Lecturer at the Victoria University Faculty of Law, with expertise in data privacy, data protection and internet law. He expressed concern at the concentration of power in the digital platforms. “This power is self-legislating. They are very loosely constrained by the laws, very loosely. Basically, they do what they want with data about people based on their own policies. So self-legislating, self-executing and in large part self-controlling.”

But the market dominance of these tech companies doesn’t only give them a perhaps unprecedented power to either ignore or negotiate with governments over the content of legislation, it also has allowed them to dominate their domestic competitors including traditional media and advertising companies. Interviewees, for example, pointed out that despite devouring much of the advertising revenue which traditionally supported New Zealand mainstream media companies, Facebook and Google resist being held accountable to the same standards of corporate responsibility regarding transparency, content moderation and taxation.

WHAT’S DIFFERENT ABOUT NZ, COMPARED TO OTHER COUNTRIES?

All of these topics have been explored in great detail globally. This research is focused on the question of how the opportunities and risks presented by digital media play out in the context of democracy in New Zealand. So one question for the interviewees was whether New Zealand was particular in anyway, and if so, how?

The most common response to this question was that New Zealand’s relatively small size gives some particularity both to the nature and style of our democracy, and to the way people use digital media.

Digital campaigner Andrew Burns suggested that New Zealand’s small size means that political debate here is potentially less toxic. “[A] lot of the debate that you do see on Twitter in New Zealand does remain quite legitimate and good natured. ... Many people know each other, so there isn’t an interest in being really rude to someone, or being too inflammatory. And people do have debates on Twitter where people will be debating in public, and then they’ll be having a little chat in private messages, behind the scenes, laughing.”

Burns, who ran the New Zealand Labour Party’s digital campaigns, thinks that political parties in New Zealand are relatively self-regulated, although that could easily change. For now, he says, the parties in New Zealand are likely to listen to the advertising standards authority, and to be aware of the political risk of doing something that would be really seen as dodgy. “I think in New Zealand it will be a lot harder for a political party to get away with doing something which most people would perceive to be really dodgy in terms of misinformation on Facebook.”

An example of this self-regulation, although by media outlets rather than by political parties, was when there was an attempt to spread disinformation about the Prime Minister's partner. In that instance, by and large, the mainstream media took the position that there was no evidence to support the allegations and were not willing to report the allegations. So the spread of the disinformation was constrained by media standards, and editors tended to err on the side of caution. As David Farrar has reported, this didn't prevent the rumours from spreading by word of mouth offline, but they were unable to spread into the mainstream online because of the constraints imposed by media.

Similarly, Farrar says, New Zealand's small size makes it relatively easy for citizens to access politicians. "We are still quite small, most people do still know their local MP. If you, I'm not sure you call it elite if you travel in the Koru club, so many people do, but you generally can run into half the cabinet in the Koru Club. I think that's nice. I think the size of New Zealand helps."

Member of Parliament Melissa Lee, agrees that politicians in New Zealand are actually much more accessible than, for example, in the United States or Korea. "[T]he very fact that I am out door-knocking and shaking hands and making sure that I communicate." And this accessibility extends to the digital space, in which many MPs are personally managing their own social media accounts and responding directly to communications from constituents. "The other day somebody [sent me a message on Korean language messaging app, KakaoTalk] and I answered back and they said, 'Oh my god I got an answer. I never expected an MP to actually respond'." Lee explained that for New Zealanders with dual residency, they might not be used to having a politician respond to a direct message like that in their other country of residence.

Lee's reference to KakaoTalk points to another feature of New Zealand's digital media landscape, which is the wide use in New Zealand of non-English language social media and messaging platforms. Chinese platform WeChat reportedly has almost 200,000 monthly active users in New Zealand. Although no official number of KakaoTalk users in New Zealand is available, one person interviewed for this report estimated that "If there are 35,000 Koreans in New Zealand then there are probably 35,000 KakaoTalk users here as well, maybe more because people like me, who are not Korean but used to live in South Korea, also use it." While these platforms and messaging apps may not be well-known to some New Zealanders, they are accurately described as giants in the global sector (WeChat is one of the world's largest standalone mobile apps with over 1 billion monthly active users) with a significant presence here in New Zealand.

Brenda Leeuwenberg is a member of the Digital Economy and Digital Inclusion Ministerial Advisory Group, and was formerly the head of Digital and Innovation at NZ on Air. She says that the focus in New Zealand's culture and policy of "making sure different aspects of our society are reflected in the stories that are told," is another unique feature of our country in the context both of stories and democracy. This is a key focus of NZ on Air's role allocating public funding to media. "[T]hey actively ensure that there is diversity across culture, ethnicity, race, gender, and everything. They actively seek of stories that tell all of our different perspectives. I think New Zealand's quite lucky like that to have that, it's pretty unusual." This public funding for a diverse range of domestic stories - whether in the form of news or drama - may help to counterbalance the impact of the multinational digital media companies on local media, although as Leeuwenberg notes, those local stories still need distribution and increasingly that includes digital.

WHAT ARE THE OPPORTUNITIES?

So given what has changed with the rise and digital media, what has stayed the same, the structural underpinnings of the major digital platforms and the unique characteristics of New Zealand's democracy - where are the biggest opportunities? The obvious and most commonly cited opportunities are in the democratisation of information, increased diversity in public discourse, more public engagement with government and democratic process, and in increased transparency and openness in government.

DEMOCRATISATION OF INFORMATION

Greater diversity of voices was one of the great opportunities promised by social media, says one participant. "That's what we were all sold, we were all sold the idea that mainstream media wouldn't be able to exclude your voice."

David Farrar, who runs a political blog, says digital media has delivered on that promise, at least to an extent. "You can become, as I have for some, a trusted source of information if you think the media haven't been giving a fair portrayal. This is really what I loved about blogging, especially in the early days, the ability to say, 'Actually, I don't think that's quite right.' That's something we now take for granted, I think, we're so used to it, but I think there are still incredible positives from that." Jordan Carter also points out that the digital media platforms all a less mediated commentary, "So it can broaden out, or offer the prospect of that wider set of perspectives being shared, and you don't have to have vast quantities of money anymore to make a public impact."

Andrew Butler notes that in some contexts, digital media offered the promise of "evading State censorship" although he notes that "one person's form of State censorship, is another person's form of the state exerting its legitimate power on behalf of the community." For example, he says, in New Zealand we have for years had State censorship. "[W]e never called it that but that's what we've had." He gives the example of limits on electoral broadcasting, which were designed to make sure the public knew "exactly who was saying what, who was paying for it and how much." The motivation behind these State limits on free speech, Butler says, were legitimate concerns about "money influencing the outcomes of public discussion."

IMPROVING PARTICIPATION IN DEMOCRACY

One of the most common themes to emerge when discussing the opportunities presented to democracy by digital media is the opportunity to increase participation in democracy. This theme includes discussions about simply increasing the number of people who take part in democratic processes like voting, submitting to Select Committees or communicating with their elected representatives. But it also covers discussion of opportunities to use digital media to make democracy more participatory, to create new and innovative ways for citizens to engage in, for example, deliberation and decision-making on issues that are relevant to their lives. It also includes opportunities to increase democratic mobilisation and activism, including the use of digital media by social movements. Importantly, it also includes a discussion about the ways in which participation in democratic processes could be made more equitable and diverse through the use of digital media.

INCREASED PARTICIPATION IN DEMOCRACY AND ACCESS TO POLITICS

Despite his concerns about social media, Privacy Commissioner John Edwards does think that digital media has “huge benefits for improving the terrible rates of [democratic] engagement that we have.”

Elyse Robèrt sees new opportunities for politics to be more accessible, and gives the example of the Prime Minister live-streaming from her office, and how that might pop up in the Facebook feed of someone who doesn't follow politics. “He doesn't read the paper, he doesn't watch the news because he spends his precious little spare time watching sport and talking to friends at the pub. He will see the Prime Minister pop-up on Facebook Live and he might have a thought about politics that would otherwise have gone unprompted.”

In a similar vein, Robèrt suggested that posting a question for a politician on a Facebook post or video might be less confronting than doing so at a town hall or Select Committee. “[I]t is more accessible and it is easier to delete if you are concerned for your safety or opinions.” In this way, she says, social media is helping to opening up the political discourse

Gwynn Compton also sees ways in which digital media has facilitated more communication between the public and politicians. For one thing, digital media is more cost effective than traditional media advertising, “you get far more bang for your buck in terms of reaching people, using digital media if you're a politician.” For citizens, on the other hand, emailing an MP or leaving a comment on their Facebook page is easier, and faster, than writing a letter. Finding a printer, buying a stamp, getting to a post office. “There were all these extra barriers to communicating, and even if they seem quite small, they're still enough to deter people from engaging.”

Digital media allows people to have a less mediated commentary on politics, says InternetNZ Chief Executive, Jordan Carter. Previously someone could send a letter to the editor of a newspaper or magazine, or call in to talkback radio. In both cases, the opportunity to be read or heard would depend on meeting the criteria applied by the editor or producer, respectively. Whereas the digital platforms allow a relatively unmediated commentary (although it's worth noting that our likelihood of seeing someone's comments is generally controlled, or curated, by algorithms).

“So it can broaden out,” says Carter, “or offer the prospect of that wider set of perspectives being shared, and you don’t have to have vast quantities of money anymore to make a public impact, with a little bit of nous on social media and with some low cost web resources, or even just using the free platforms alone, you can have a perspective and a debate that you couldn’t have before.”

Anna Rawhiti-Connell, a digital and social media specialist, agrees that the opportunity of digital is for citizens to be a lot closer to decision making bodies, and has also seen the rise of consumer and shareholder activism through social media.

MORE PARTICIPATORY DEMOCRACY

Dave Moskovitz agrees, and sees opportunities for a more participatory democracy in which digital technology gives people new ways to get involved in helping to formulate and decide on the issues that they care about the most. He acknowledges that we don’t yet have a good way of doing this, but he is hopeful. “I’m confident that as a society we will develop this. That there is a way to help people exhibit emergent behaviour, around coming together on issues that are important to them in a considered way.”

Barrister and law and technology researcher, Dr James Every-Palmer QC, agrees that changes and developments in digital technology have opened up the possibility of more participatory democracy. “There’s no technical reason why we can’t have online referenda to directly decide some category of issues, rather than leaving it to elected representatives ... It’s exciting. It could be very empowering.”

GREATER DIVERSITY OF ENGAGEMENT AND EQUALITY OF ACCESS

Joel Colón-Ríos is an Associate Professor in the School of Law at Victoria University of Wellington with an interest in constitutional law. He described the role of digital media in enabling independent political candidates to build popular campaigns outside of the main political party system. If independent candidates are able to use Facebook, Twitter or Instagram really effectively, he says, they can become quite popular “and sometimes will even win elections.” He’s quick to note that this isn’t necessarily a good or a bad thing. “It could be used by very good people, but also by people that have [what some may call populist] policies that ... attract the attention of large parts of the population, because they are able to very quickly and effectively disseminate a message containing precisely the things that people want to hear.”

Chlöe Swarbrick, former Auckland mayoral candidate and Green MP says she wouldn’t be in Parliament today if it wasn’t for the internet and social media. “I was completely unknown, had no money, no institutional connections. Had no idea about the establishment, *per se*, but naïvely, and idealistically fumbled with a bunch of friends through all the processes that we found out about on Google, and then created ways to engage with people on the internet.” Getting people engaged online led to getting those people along to physical meetings, and from there she and her small campaign were able to grow momentum for the mayoral campaign. Which, in turn, led to her being offered the opportunity to stand as a Green Party candidate and being elected to Parliament in 2017.

Professor Paul Spoonley uses Swarbrick's campaign as an example of how digital media has become a central part of how young people, in particular, engage with democracy. "[In the] political campaigns of people like Chlöe Swarbrick. [Digital media] is not simply an add-on to what they currently do in terms of communicating. It's at the very core, and they're beginning to express a quite different approach to political debate and political campaigning, which I think is quite a significant shift compared to what we've seen in the past."

Once those independent politicians, or politicians from minor parties, are in Parliament, digital media continues to offer them opportunities to communicate more widely with their constituents than the traditional structures of democracy, or for that matter mainstream media, might allow. Former Māori Party MP Marama Fox found that the small number of seats held by her party in Parliament meant she had very limited opportunities to pose formal questions in the House, but social media gave her alternative channels to have her say on the issues of the day. "I was only able to ask one question every eight days, and two supplementary questions a week. So I had to sit there, filtered, and silent and muted, and that just pissed me off. So I would get on social media, and engage people there to get my say out there."

INCREASED OPPORTUNITIES FOR ACTIVISM AND MOBILISATION

The Arab Spring is a commonly cited example of the role of digital media in democracy. Joel Colón-Ríos explains during the Arab Spring, people used digital media to mobilise and organise, "to say 'Go to the public square, tomorrow at 9pm, we'll have a huge protest there.'"

Sandra Dickson gave an example from closer to home, in which students in Wellington responded to harmful comments made online by male students at Wellington College. Other students used social media to mobilise their peers from across the city to gather outside Parliament to call for better sexuality education in school. Support grew quickly, she says, leading to 400-odd young people marching on Parliament. At the same time, two young women from Wellington schools starting an online petition to call for more consistent sexual consent education across all schools, which gathered thousands of signatures and was presented to MPs from all major parties.

As this example illustrates, young people are often cited as a group who are assumed to be confident users of digital media, and therefore likely to benefit from opportunities to engage in democracy digitally. But children are often left out of discussions about democracy, and child rights advocate, Sarah Morris, from the Office of the Children's Commissioner says that needs to change. Children have a right to participate in democratic processes, even before they have the right to vote, says Morris, and digital media can create opportunities for that to happen. She gives an example of a project from her time at UNICEF in which children in India collaborated with Google to map the slums where they lived and identify the gaps in being able to access clean water. "Then they took that data to local government and used it as a way to advocate. Those kinds of examples of children engaging in digital spaces for greater participation in democracy are really inspiring and useful."

Sandra Dickson also sees opportunities for greater democratic engagement from minority

and marginalised groups. She says access to digital media has enabled trans and gender diverse people, for example, to write about their experiences and to connect and organise with each other. This ability to communicate and organise online has, she says, led to the needs of trans and gender diverse people being more well-known and understood than in the past, and meant there are additional avenues of support for trans and gender diverse people.

“[For] communities that are relatively small, but with quite serious discrimination issues to resolve, that’s magnificent, because it allows people to connect to one another, internationally as well as locally, it allows a voice to be amplified more safely. [Y]es, we absolutely still see violence towards trans people, I’m not saying we don’t, but I think if the only ways trans folk were able to raise their concerns were in person, we would have seen a lot more physical, life-threatening violence towards trans people.”

Sarah Morris also sees opportunities for more democratic participation by children living with disabilities. Digital spaces provide ‘mind-blowing’ potential to participation and democratic engagement by children with physical and intellectual disabilities and neurodiversity, says Morris, “that’s a really amazing area to look at.”

TRANSPARENCY IN GOVERNMENT

Another major theme in terms of opportunities arising from digital media is the potential for greater government transparency. As Jordan Carter explains, digital technology has made it possible to “render more of the decision making processes of government transparent.” He says it “hasn’t been terribly well used” but in principle, the policy making process could be digitised and made much more widely available.

Thinking about the role of government as “an enabler and protector of rights and liberties”, Pia Andrews argues that government should be as open as possible. “If the rules were all available as code [people could] test whether the policies are being applied the right way,” which, she argues, would create more transparency and trust. Transparency enables traceability, she says, which becomes even more critical as more government decisions are automated through AI. “[I]t is critical that we have traceability and accountability and transparency back to how decisions are being made.”

One of the ways that digital media can help increase the transparency of government is through increased citizen engagement with democratic processes that hold politicians and civil servants accountable. Dr Every-Palmer gives the example of campaigning organisation ActionStation, which uses digital media to enable people to engage democratically on policy issues that matter to them. “[They] seem to be a poster child for the ability of new technologies to make a difference in terms of mobilising or allowing a group to form over a particular issue, and ensuring that their voice is heard. The facilitation of interest groups is another consequence of the lower costs of communication and coordination in the online world. It is exactly how many people twenty years ago imagined the internet would work.”

Sandra Dickson also sees opportunities to use digital media to create public pressure on social issues that lots of people care about, “and being able to bring those to the surface in a way that is outside of the parliamentary processes, but which creates pressure on decision-makers.”

Social media also makes it harder for decision making bodies “to sweep things under the rug,” says Anna Rawhiti-Connell. Digital proximity “creates a greater awareness of the things that matter to people, for organisations and decision making bodies. I think that creates a greater transparency.”

As a counterpoint, Marama Fox insists that meeting face-to-face is still essential for politicians and for political discourse generally. “The keyboard warrior disappears and you’re more willing to have a calm conversation about something, rather than an argumentative one.” Fox also points out that the combative tone of much digital discussion of politics will put off some people, and specifically people from certain cultural backgrounds. “[Some] Māori, like my husband, hates me arguing around the kitchen table with my sons. Fox says that because she was educated in a Pākehā cultural context, she is more comfortable with that sort of confrontational discussion. “He calls it arguing. I call it a debate.”

Scanning the landscape, James Every-Palmer says “there are some areas in which new technologies have made a positive difference; in access to news, and initiatives like ActionStation. But overall it doesn’t feel that we’re living in a more healthy democratic environment than we were two decades ago, when you look at fake news, filter bubbles and outside interference in elections, I think the reverse might be true.”

A CASE STUDY IN TRANSPARENCY: MPS ON TWITTER

Citizen AI is a charitable company whose mission is to research, develop and promote AI systems for public benefit. In 2018, they became interested in how MPs interacted with citizens on Twitter, and whether it was possible to map who had influence over MPs according to their Twitter interactions.

Citizen AI founders Matthew Bartlett and Geoffrey Roberts were interested in the way that you could tell quite a lot about someone by the people they follow on social media, and thought “maybe it would be interesting to look at who people actually like, or what things people like on Twitter as a way of actually categorising them.” So they created a map of New Zealand Members of Parliament on Twitter, taking data available from Twitter’s API to map out the last twenty tweets that each MP had liked.

Asked what they had noticed about their map, the Citizen AI founders said they were more interested in seeing what other people had noticed. But what they did notice was the polarisation, by which they refer to the pattern in which most MPs did not interact with “the other side”, and the fact that MPs largely favourite each other’s tweets, rather than those of members of the public. Bartlett says, “[t]here are lots of MP’s who will retweet and sneer at the other side and that’s why we excluded retweets from the first cut.” He also noticed that “[t]he National Party seemed really tight” in their use of Twitter, and wondered if they might have a policy that MPs should, for example, retweet the party leader.

The project was, by Bartlett's own admission, limited. "It began as a way of exploring a new programming language. But the project could be built on to provide a much more sophisticated picture of who our politicians are paying attention to, by for instance exploring shifting political alliances as they develop over time."¹

But it provides an interesting case study in ways that specific slices of the intersection between social media and democracy could be represented in accessible, visual and transparent ways. Twitter was a good example to choose for this project partly because of the way politicians have used it to engage with each other, with journalists and with a particularly politically engaged section of the public. But it was also a good platform to map because its API made more data accessible publicly than other platforms. The flipside of this, says Bartlett, was that the project "felt a little bit invasive in a way — it's strange that Twitter has this 'default public' aspect, and that stuff is just out there."

GREATER ACCESS TO INFORMATION

Access to information, particularly information relevant to matters of public interest, is an essential element in a functioning democracy. A number of interviewees said that digital media had improved access to information, including information both from and about government.

John Hancock, Senior Legal Advisor at the New Zealand Human Rights Commission, says one of the positive aspects of the rise of digital technology has been that the information of government and democratic processes is more accessible than it used to be. "A member of the public can access pretty much any government policy online, once cabinet papers have been released, they're there. Legislation is there, decisions of the courts are there. It has opened up the institutions a lot, as well. They've become a lot more accessible than they used to be, through digital media"

Jordan Carter agrees that one of the opportunities of digital media is that "information can be more widely promulgated by decision makers, by political candidates, by people advocating for or against an idea or a position or a policy." Compared with the costs of print or broadcast media, the costs of distributing information online are very low. This not only allows citizens to have their say, but also reduces the costs of publication for institutions. Digital media also opens up a wider range of channels and diversity of forms through which public institutions can communicate with a broader range of people with a stake in their work. "You can provide the 30 second sound bite, the three sentence summary, or the 280 character tweet, but you can also link it to more detail."

Child rights advocate, Sarah Morris points out that under the United Nations Convention on the Rights of the Child, children have a right to access reliable information from the media. This is essential to their enjoyment of their right to participation and to freedom of expression, both also protected in the Convention. Digital media plays a large role in the ways that children access information, and while this provides some really positive opportunities to ensure children enjoy their right to access information, Morris points out that it is also really important to ensure that children are able to access reliable and accurate information, which can be a challenge when using digital media.

Thomas Beagle, Chairman of the NZ Council for Civil Liberties and co-founder of lobby group Tech Liberty, says that one of the features of digital media that makes it a powerful

1. The project is visible at <http://mps.citizenai.nz>, and the open source code is available at <https://github.com/Citizen-AI/twitter-network>.

space not only for accessing information, but for actually learning, is that it gives a platform for people to share their personal experiences. “[P]eople are talking about their own experiences ... If you’re thinking about transgender issues, for example it’s been really interesting having some trans people speaking up and saying ‘look this is how it is, this is what that means.’ That’s very very powerful, as a means of communicating and persuading people.”

A MESSY PICTURE

Joy Liddicoat is a law and technology expert, currently researching human rights and artificial intelligence at the University of Otago. Formerly a Human Rights Commissioner, Assistant Privacy Commissioner and currently Vice President of InternetNZ, she has followed the growth of digital media closely, with an eye to its impact on human rights. Like many others interviewed for this project, she says the early promise of digital media was diversity - “a thousand flowers blooming of lots of different opportunities for democratic engagement.” And, she says, digital media has created new spaces for democratic engagement. “[O]nline was this space for democratic engagement, to express political views, to organise, to protest.”

Liddicoat does see progress arising from digital media. “[W]e see innovation, we see creativity, we see breaking of media monopolies, so we have more choice about how we get informed.” But on the other hand, she says, we are more and more dependent on the people who are curating that information for us, and directing our attention. The same, she thinks, is true in government. “[T]here’s so much government dependency on digital media to engage in democratic consultation, but it’s impossible to follow how many. Most big government agencies couldn’t even tell you how many things they were consulting on at any one time.”

Overall, she says, the picture is quite messy, in terms of how democracy as a form is evolving under the influence of digital media. “[I]t’s got all these new ways to participate, all these new channels for participation. At the same time, it’s getting harder to curate and access that content online, and also critique it. So it’s a messy space to talk about risks and opportunities, because the whole landscape is so complicated and moving.”

In general, the interviewees in this research tended to agree with Liddicoat. They articulated many opportunities for democracy arising from digital media, but tended to qualify those statements immediately by setting out some of the risks and problems.

WHAT ARE THE RISKS AND PROBLEMS?

DIGITAL EXCLUSION

The first challenge that arises when considering the rise of digital media on democracy is the ongoing digital divides that prevent some people from being able to access and participate in digital spaces. While the number of people who are unable to access the internet in New Zealand is small, recent research has highlighted that they are more likely to also be experiencing other forms of social exclusion. This means that increasing reliance on digital media to enable democratic participation - whether that involves staying informed about issues of public interest, communicating with elected representatives or having your say - is likely to disadvantage people who are already excluded in other ways.

In 2018, The Workshop published *Out of the Maze*, a report based on research jointly commissioned by InternetNZ and the Vodafone Foundation to uncover the lived experience of people who are digitally excluded in New Zealand. The findings of this research are relevant to this report because they highlight the connections between digital exclusion and social exclusion, both of which undermine people's ability to engage democratically.

Out of the Maze affirmed what previous research had shown. Having access to affordable and accessible digital devices and services at a convenient time and place, as well as the motivation, skills, and trust to use the Internet to pursue and realise meaningful social and economic outcomes, is essential to social inclusion.

Participants reported that losing access can have a disproportionately harmful impact on people who are already experiencing social exclusion in other ways. Becoming disconnected can have a devastating impact on people going through times of particular vulnerability or instability. A woman leaving a violent relationship, for example, or a young person leaving home without any family support.

The barriers people identified to Internet access fell into six broad categories:

- > cost
- > physical access (including infrastructure, accessible spaces and adaptive devices for people with disabilities)
- > motivation (including the motivation of parents and guardians)
- > trust and safety
- > skills
- > capacity.

While most of this was consistent with research in other countries, Out of the Maze gave a more nuanced and detailed picture of how a variety of different social and economic factors can act together to impact a person's digital inclusion at different times throughout their life, and how becoming disconnected can, in turn, exacerbate existing problems.

On balance, it seems clear that digital inclusion is essential to the social inclusion, and therefore the democratic participation, of the vast majority of New Zealanders, and especially young people. Having the skill to navigate the Internet both safely and with confidence is essential to genuine digital and democratic inclusion.

Neil Melhuish, Director of Policy at Netsafe, says that factors related to digital exclusion could be undermining the ability of some New Zealanders to engage democratically, and emphasises that the barriers that exclude people from digital spaces includes intimidation or harassment. "There are a myriad of factors which might influence whether people go online and participate in a political debate. For example, digital access - including connectivity and skills could be an issue - or perhaps aspects of the online environment such as a platform's functions or how discussion is facilitated. If for whatever reason you feel intimidated to be in a particular environment, or to have your own opinion, you are less likely to engage".

Melhuish sees a connection, specifically, between online abuse and access to digital opportunity because when people don't feel safe online, they are less likely to engage in certain kinds of conversations online. "There's evidence from Netsafe's research to suggest that people alter their behaviour around technology and the impact of their experiences on the way it makes them feel, if they're in spaces where it makes them feel vulnerable or under attack." He also notes how this online experience can impact on offline behaviours too.

Another aspect of digital divides, says Pia Andrews, is the simple problem that people are busy, too busy to have the capacity to engage meaningfully with many democratic processes - digitally facilitated or not. "The problem," she says, "is that people are working a 50, 60, 70-hour work week, then at home there are stresses on school, on homework, on all of the stuff that it takes being a parent today, which is more than it ever has been." So no-one has the time to participate, and yet public servants are looking at democratic participation as a 'supply chain' problem. "They're saying 'what we'll do is we'll build a mountain, and anyone's welcome to come climb it,'" and then you have a thousand mountains popped up. Firstly, how would you even know they exist? And secondly how would you ever climb all of them even if you were interested?"

Governments, with good intent says Andrews, have focussed entirely on the supply issue of participation in democracy, “without realising that we actually need to rethink it because those approaches are linear and don’t actually scale to the exponential changes that are happening. And the complexity and change and speed of change affecting people’s lives is only going to get faster, and faster, and faster.” What’s needed, she says, are solutions that look both at the demand side of the equation - what can be done to reduce the general life load on busy citizens - and on designing nonlinear solutions that acknowledge and respond to the exponential nature of the changes brought about by digital media.

MONOPOLIES AND LACK OF COMPETITION

Another structural challenge posed to democracy by the rise of the digital giants is the way in which their scale and relative power has too often allowed them to operate outside of the reach of local regulation including antitrust regulation. This undermines the functioning of competitive economy, one of the core features of a modern liberal democracy. In general, it gives these multinationals unfair advantages over local companies and particular it has created unfair disadvantages for domestic media, advertising and digital media companies.

Privacy Commissioner John Edwards reflected that antitrust law hasn’t yet worked out how to deal with the global platform monopolies and that there needs to be an attempt to bring some uniformity to the international legal approach to this.

“We had this issue with Facebook earlier in the year. We said ‘well you’re here, you’ve got two and a half million NZ’ers data that you’re hoovering up, you’re collecting advertising revenue from NZ businesses, you’re telling people in New Zealand to go to this cafe that their friend went to.’ There is very much a presence in NZ that is touching on many aspects of the NZ economy, but purportedly without having to be troubled by the legal framework that all the other competing businesses are acting in.”

Ganesh Nana is Chief Economist at BERL (Business and Economic Research Limited), with over 30 years of experience in the field of economics. As a general principle, he says, businesses have a right to operate in a country or in a jurisdiction only insofar that they abide by the regulations in that country. And in principle, the rules of engagement for those multinational companies should reflect the rules that are imposed on all other enterprises in New Zealand.

“[O]ne of the core rules in terms of free trade agreements and cross border economic activity is that we are not allowed to discriminate between businesses that are offshore versus businesses that are in New Zealand. So the principle would be that the rules we want to impose on them are the rules that we would impose on any organisation operating in New Zealand or any business operating in New Zealand, whether they happen to be owned by New Zealand or foreign owned.”

One impact of this is where we’ve seen advertising dollars moving to the big platforms because they can offer advertisers things that their domestic competitors can’t, because the platforms have simply declared, in the case of Facebook, for example, that they are not constrained by our privacy laws. So, there is a fundamental unfairness there between

the standards that domestic media companies, for example, are held to and the standards that these platforms are held to.

As Nana puts it, “we would call that a discrimination. That is clearly unleveling or tilting the playing field against domestic firms and that would go against that principle of a level playing field.”

Beyond ensuring multinational companies are required to abide by the same standards as their domestic competitors, which is difficult enough, it is particularly difficult to create meaningful competition in a market where the success of the product is based on everyone you know using the same product. As Nat Dudley says, “you have this whole online culture that’s predicated on the idea that in order for a platform to be successful they basically have to be a monopoly in their area. There’s a huge inertia once you have a strong player in that market.”

Dudley points out that there is a unique social cultural aspect to the inertia in this case because these platforms are not just the places where everyone buys their books or sells their cars. They are the spaces where we keep up with local, national and international news, where we talk to our friends and organise out social gatherings. Which makes them particularly hard to move away from. “And you have this monopoly that’s not transparent. You can’t see how they’re deciding whose ads we see, or who sees a message from Facebook encouraging people to vote.”

Another structural factor which makes these platforms fundamentally different to what came before them, says Dudley, is the role of venture capital funding.

“All these big platforms are venture capitalist funded. And all of them bring that motivation. They’re not regular organisations looking for slow sustainable profit, they’re looking for to maximise the massive investment they got, which is all about profit and scalable cost for venture capitalists. They’re the bigger problem in our industry, and what drives a lot of the aggressive strategies by these platforms, which they’ve invested millions of dollars in, and they want the returns back. So you’ve got this other shadow party sitting behind them which makes it difficult to see how they’d negotiate in good faith, because they’re not the power broker in that situation, they answer to their board and investors beyond anyone else.”

IMPACT ON PUBLIC- INTEREST MEDIA

One of the impacts of the virtual monopoly of digital media platforms has been to undermine the funding base, and therefore viability, of traditional forms of public interest media. As many interviewees observed, while Facebook and Google have taken the lion's share of advertising dollars away from traditional media, they do not themselves create the kind of expensive, essential public interest journalism that is essential to a functioning democracy.

The ability of traditional media to compete with the digital platforms for advertising spend has been undermined by double standards of accountability. As media lawyer Andrew Butler explains, "the whole point of what we've been fighting for, for 200 years, is public purpose. Mainstream media again has social license, and with the social license comes responsibility. You can't just say whatever you want to say. If you're going to be able to influence public opinion, then you've got to do it in a fair responsible way." That's what the Broadcasting Standards Authority and Press Council/Online Media Standards Authority are all about. But these same standards haven't typically applied to the digital platforms, putting them in an unfair position as competitors to traditional media outlets.

The irony of projects like the Google News Initiative - "an effort to work with the news industry to help journalism thrive in the digital age" - is that Google are simultaneously dismantling the foundations of public interest journalism while, with the other hand. "Google cares deeply about journalism," says Google CEO Sundar Pichai. "We believe in spreading knowledge to make life better for everyone. It's at the heart of Google's mission. It's the mission of publishers and journalists. Put simply, our futures are tied together."

This may well represent the genuine beliefs of many at Google, but it is hard to reconcile with a business model which has had such a devastating effect on the traditional funding sources for journalism. Much of traditional media's revenue, says Ian Apperley, came from what he calls the 'buy/sell/swap ads' that were listed in the back of the paper. That revenue began to diminish with the emergence of online trading sites like TradeMe. Then when Facebook started to sell advertising to everyone from the local restaurant to political campaigns, and allow advertisers to limit their spend to a carefully targeted audience, the revenue largely vanished.

So, says Apperley, traditional media "are going to die." Unless they adapt, presumably. And he concedes there are "some reasonable models around", but because most of the new business models that offer hope for the future of journalism involve subscriptions and paywalls, he says, "they bring back an access problem." Using the National Business Review as an example, Apperley says that you can get really interesting business stories and information if you have a subscription. But only if you can afford it, and most people can't afford to do that.

Some people, like historian Paul Moon, argue that if people aren't willing to pay for good quality public interest journalism, that means they don't want it. Media commentator Colin Peacock, on the other hand, says that research has shown that people do value public interest content, even if they don't watch it. "People were saying, even if they didn't watch documentaries, they were happy that TVNZ was making and screening them." There is a suggestion in this, says Peacock, that people understand the value of public interest journalism in a health democracy and want it to exist, even if they don't choose to sit down and watch a documentary on steel frames in buildings at the end of a long work day.

MISINFORMATION, DISINFORMATION AND MALINFORMATION

Economist Ganesh Nana describes this as an externality, i.e. people want public interest journalism separate to the market. “From an economic perspective, I think there is an externality benefit from a strong vibrant journalism sector within the country, with that benefit being the maintenance and strengthening of democratic processes and the flow of a generation of respected information.” So if there is a market failure happening, because for example all the marketing dollars are going to organisations like Facebook that are not generating news, [digital media platforms such as Facebook], Nana explains, the community, society, democracy is being harmed. “There is a public good benefit of having an informed community who can then contribute and connect and participate in a vibrant democracy.”

Peacock agrees there is a threat to democracy, but sees the greatest threat at the local level. “The threat is in smaller communities, and that local side of democracy. Are people getting what they need? Are they going to be exposed to it? That’s where I see the threats, and the opportunities.” Even in a publicly funded broadcaster like Radio New Zealand, Peacock says, the impetus is not to provide better local news.

Another democratic risk arising from the erosion of funding for traditional media has been the erosion of trust in mainstream media. As media outlets fight to compete with the digital platform companies for advertising dollars, they chase clicks and impressions with exaggerated headlines, alarmist stories and celebrity ‘news’.

It’s a survival strategy that risks undermining public trust in the media, which is a serious threat to democracy according to Nat Dudley. Not least because it undermines their ability to counter misinformation and disinformation. “So even if the media could critique it and say actually this is factually incorrect or this is really awful the likelihood of it changing the opinion of those people is pretty slim. Removing people’s faith in the media as a place they can find some form of truth is a really, really destructive thing. Because where do you go and find critique of this if not the media?”

When we started scoping this research, and realised the scope might be too big for one project, some people suggested that we focus on what they saw as the biggest threat to democracy arising from digital media: fake news.

It’s a topic that has fallen, ironically, prey to its own fate, with endless misleading allegations circulating about what is and is not fake news. For the purpose of this research, we needed to start by exploring a working definition of fake news and quickly decided that the term was simultaneously too vague and too corrupted to be useful for our purposes. Instead we use the definition developed in Information Disorder, a 2017 report by Dr Claire Wardle and Hossein Derakhshan who introduced a new conceptual framework for examining what was often referred to as ‘fake news’.

Wardle and Derakhshan identified three different types: mis-information, dis-information and mal-information. Using the dimensions of harm and falseness, they describe the differences between these three types of information:

1. Misinformation is when false information is shared, but no harm is meant.
2. Disinformation is when false information is knowingly shared to cause harm.
3. Malinformation is when genuine information is shared to cause harm, often by moving information designed to stay private into the public sphere.

In this report, we use these three terms with these meanings, while also recognising further complexity within and between these terms, including examples where factually correct information is presented out of context, or in a misleading context, so as to leave a false impression on the reader. This can be done with or without the intention to harm. While intent to harm on the part of the producer or distributor of the content is relevant to understanding the risks to democracy, so is the harmful impact of the information, whatever the intention may have been.

A regularly cited example of harmful misinformation which is being widely shared without the intent to harm is anti-vaccination content. Many of the people sharing, and perhaps even the people producing, misleading information about the risks of vaccination are doing so without any intent to harm. They genuinely believe that their content is accurate, and therefore share it with the intent to help people. This good faith belief in the accuracy of false information does not, however, mitigate against the harm that is done by creating and distributing that information. This example helps illustrate the impact that misinformation, through digital media platforms, can have on how well-informed citizens are when they participate in democratic processes like voting or submitting to Select Committees.

Because of the overlap between misinformation, disinformation and malinformation, this section will move between the three while examining interviewee' views on the level of harm and potential risk each pose to democracy in New Zealand.

MALINFORMATION

Perhaps the most well-known recent example of malinformation in New Zealand politics is set out in Nicky Hager's book *Dirty Politics*. In 2014, Nicky Hager was leaked a large number of email and online conversations from Cameron Slater's Whale Oil blog. Many of these were between Slater and his allies on the hard right of New Zealand politics. Amongst the leaked communications were many with the Prime Minister's office and other Cabinet ministers. They revealed use of digital media, including through Slater's own blog, to spread both disinformation and malinformation, with the goal of undermining not only politicians on the left and their staff, but also public servants, journalists and even academics whose work presented challenges to the policies of the National Government at the time.

Hager explains how these practices undermined core elements of democracy in New Zealand. "When Slater was attacking the left wing, a press secretary of an MP. Or attacking public health professionals. They knew exactly what they were doing. You get in someone's head, you scare them, and meanwhile, other people are thinking, 'maybe I won't stick my head up on alcohol,' or whatever it is. It's about trying to take out your

opponents. Not trying to win the point but trying to remove the activity and the voices of your opponents. It's about as bad as you get, democratically."

Along with the direct impact of silencing people whose work raised challenges to the politics and policies of the government at the time, Hager says these practices undermine public confidence in politics generally, thereby undermining public participation in democratic processes. Citizens are left thinking that all politics are dirty, Hager says, "I don't like any of them. They're all the same. They don't like politicians." This effect, he says, is not accidental.

"[T]here are people who are specialists in that attack delivery as well, it's not an accidental thing. It's people who don't believe in democracy. I don't want to be apocalyptic, but those are the things that scare me the most."

While attack politics existed long before digital media, the rise of blogs and social media made it easier for bad faith actors like Cameron Slater to publish and distribute disinformation and malinformation.

DISINFORMATION

Beyond the examples in *Dirty Politics*, most interviewees struggled to think of clear examples of intentional disinformation in New Zealand. This could be because, as some interviewees suggested, there is less of it here than in other jurisdictions. Or it could be because, since *Dirty Politics*, practitioners of attack politics have adopted more subtle approaches that are harder to identify.

Tech journalist Ian Apperley notes that, compared to the United States, New Zealand hasn't seen much in the way of long form disinformation, "We haven't quite seen the InfoWars-type of stories, where it's this long running narrative of more, and more information that's incorrect, or misleading, or any of those categories."

Paul Spoonley also observes less inaccurate material being published in New Zealand, including on digital media, "we've got very little in this country that monitors and regulates material apart from groups like NetSafe." Andrew Burns agrees, "[I]n New Zealand people haven't actively started creating meme-type pages to influence people. All the political parties have funny meme pages which are run probably by random activists. But I don't think we've seen that uprising of someone deliberately, or even accidentally, setting out to start influencing people."

'DEEPPAKES'

Some interviewees noted, however, that deliberate disinformation created outside of New Zealand could still be having an impact on our democracy. Well-documented use of disinformation to undermine Hillary Clinton in the lead up to the US presidential election in 2016, for example, could contribute to the kind of overall sense that politicians can't be trusted, which Hager refers to in *Dirty Politics*.

The creation and spread of 'deepfakes', whether created in New Zealand or elsewhere, is also a cause for concern to some interviewees. Tom Barraclough and Curtis Barnes (Brainbox Ltd) have prepared a legal research report into "deepfakes", an example of a

wider family of technologies known as “synthetic media”. Synthetic media of all kinds will pose a growing issue because a wider range of people will gain the capacity to make it look or sound like something happened when it didn’t happen.

The potential of synthetic media such as deepfakes can be scary, but many of the harms they produce are the kinds of harms already dealt with by law. Harms can come from being too sceptical as well as too trusting of audiovisual content. Harms can also come from allowing state and private actors to unjustifiably restrict civil society’s ability to create and disseminate audiovisual information.

Deep fakes are concerning to many because they are extremely convincing. As Andrew Butler explains, most of us have been taught that our eyes do not deceive us, “if you see something, it’s very compelling.”

Brenda Leeuwenberg agrees, and says it’s also getting harder to train children to distinguish good information from bad. “I used to say it’s really important that schools teach kids how to verify information, how to verify a source, how to identify fact from fiction. I don’t know that it’s that easy anymore.”

And it’s not only the deepfakes that are harder for people to recognise, says Andrew Burns, many people find it difficult to distinguish advertising from editorial content. “I want people to be really aware, from a democratic perspective, of what is an ad, and what isn’t.” But Burns says that is just not easy for many people, “because it’s all just a stream of stuff.” Advertisements made for corporate clients might be easier to recognise, he says, “there’s a production value, there’s the way the copy is driven.” But what is much harder to spot is an ad that has been created to look like user generated content. “Looking like a crappy post that someone has put in a group, and one of your friends commented on it, or something like that, and it just happens to come from a page that sounds like something you might agree with, and then it slowly radicalises people.”

MISINFORMATION

Most interviewees saw more examples in New Zealand of misinformation - where false or misleading information was being shared by people without the intention to harm. The common examples given of misinformation were anti-vaccination and anti-1080 content, much of which is spread, according to interviewees, by people who are convinced that the content is accurate and that by sharing it they will help reduce, rather than cause harm.

One of the structural changes, led by digital media, which has created the conditions for so much false and misleading content, says author Nicky Hager is the deregulated nature of the sources of content which is being spread by and through digital platforms.

“[Y]ou have deregulated sources. Unregulated, irresponsible sources. Self-interested sources, which can just mainline stuff into the public sphere. And then magnify it through comments sections, and fake commenters. Where no-one knows who they are, and all the rest of that. It’s really easy. [I]t’s just, it’s made for manipulation.”

Member of Parliament Golriz Ghahraman has herself been the target of misinformation and has noticed an interplay between digital media and mainstream media reporting in the spread of misinformation. “Basically there’s a really clear pattern where information appears on social media, with none of the gatekeepers of editors or checks or anything,

but then it gets reported in mainstream media that it was said in social media.” Ghahraman points out that when the media report on misinformation, they are not claiming the information is accurate. They are reporting that someone on digital media has made these claims “[B]ut I just think people don’t distinguish.”

Sandra Dickson says one of the areas in which she sees a lot of misinformation, some of which could be considered disinformation, is in the misreadings Pākehā have about New Zealand history, including false accounts of the history of the Māori.

“Any issue that’s about fundamentally shifting power,” is ripe for misinformation, says Dickson. She gives the example of family and sexual violence, a field she has worked in for 28 years. She can say something about what the evidence shows about, for example the causes, patterns and incidence of family and sexual violence against women, “And I can have a man say, ‘oh but I think men can be victims too,’ and we will have the same validity of platform.” So the opinions of people who are misinformed, or who have some information without the full context that makes sense of that information, can be given a kind of equal standing on digital media, which, Dickson says, can lead to further public confusion and misunderstanding about matters of public interest.

In the context of these imbalances in authority, Dickson says that public discussions of some issues are being presented as balanced when a false equivalence has been given to misinformation, alongside more accurate information. Conversations between very opposing views are presented as though they are democratic, says Dickson, “even when this view has this much evidence, and this opposing view has none.” This leads to an overall misleading construction of the conversation.

Misinformation is more convincing when it feeds into an existing belief, says David Farrar. ““You look for stuff that will validate what is beneficial to you. [For] most of the anti-1080 people it’s quite simple. They like to hunt, they worry about the impact on their hunting, they don’t really care about the impact on native bush, so they’re going to do what they can to stop 1080 because they like to hunt, you know?”

Echoing the findings of a growing body of research on human cognitive responses to information which is alarming, Andrew Geddis notes that people are particularly susceptible to misinformation if it engages fear or anxiety. “[P]eople for some reason – whether just in our society or some sort of deeper human nature thing – just seem attracted to information that raises fear in them and causes the alarm response.”

Andrew Burns notes that digital content overload can be used as a tactic to spread misinformation. A small number of people can effectively ‘spam the internet,’ says Burns, turning ‘regular people’ off from participating in public discussions about the issue. He gives the example of anti-1080 protestors who adopted the tactic of posting hundreds of questions and comments about 1080 on every Facebook live hosted by the Prime Minister, no matter what the topic of the video. “[T]here are people who are really angry about these things,” says Burns, “but its stifling, and it’s quite a small number of people. It’s a really small number of people, in an organised effort to spam the internet with anti-1080 information.”

Another tactic used to spread misinformation, says Burns, is ‘chaotic misinformation’ which is a term he uses to describe social media posts which misrepresented the Prime Minister’s view of abortion law reform, “to mean that babies that might have a

disability will be aborted. It was very targeted, and it was stretching the truth beyond belief. It was like a chaotic misinformation.” He says he thinks the people spreading the information genuinely cared about the issue and were genuinely misinformed, rather than intentionally misrepresenting the Prime Minister’s position. Burns says it was raised with Facebook, and they said the posts didn’t breach their terms of service. “[I]t was quite a marginal case, and there wasn’t really much we could do about it, because it was someone’s genuinely held belief. [I]t wasn’t defamatory or anything like that. It was just one of those edge cases where it wasn’t true, but it was a genuinely held belief.”

EROSION OF TRUST IN INFORMATION

One of the general concerns raised by interviewees about the impact of misinformation on democracy was the risk that it will erode public trust in reliable information. Marcin Betkier says the risk of exposure to misinformation is not only that people will believe it, but that they will stop recognising and believing genuine experts. Because there is so much information around, says Betkier, and because so much of it is unreliable, “we don’t have the reliance on authority that we used to have, whether academic or any other kind. People look at YouTube, for example, where authority is measured in likes and clicks, and this undermines the authority of people who have some knowledge in topics.”

Chloe Swarbrick notes that this erosion of trust in information has led to a context in which the perception of authenticity is more important than accuracy. “Trump is such a fascinating and terrifying example, because despite being emblematic of everything that was wrong in the system, he was able to say that he was anti-establishment.” One of the factors that made his message compelling, says Swarbrick, was “the perception of authenticity he had, at least to his supporters.”

POLITICAL MANIPULATION INCLUDING FOREIGN INTERFERENCE

When people talk about the use of digital media for political manipulation, they use that terms to cover a wide set of activities ranging from interference in elections by hostile foreign states through to digital microtargeting of political advertisements in ways that make it difficult for anyone to have a single, clear view of the policies and political rhetoric being presented to each specific segment of the target audience. This range of activities spans from the highly illegal - although difficult to trace - through to actions which are perfectly legal but which arguably erode the kind of shared understanding of public issues which is considered to be essential to an effective democracy.

CYBERSECURITY OF GOVERNMENT AND SECURITY OF ELECTIONS

When you talk about digital media and democracy, many people will immediately associate the topic with the risk of foreign interference in elections. “We should have measures to protect our elections, to protect our critical infrastructure,” says Marcin Betkier. He says there is work going on by the Government Communications Security Bureau and the Department of Prime Minister and Cabinet, “there are some programs in place to secure the infrastructure, we just need those measures to make sure that we are sovereign in this respect”

As founder and owner of ZX Security Limited, Simon Howard has worked on information security for clients in the banking, energy and government sectors. He and his team are experts in testing the security of, and finding vulnerabilities in, large digital networks and systems. The New Zealand government, Howard says, need to be proactive and vigilant about ensuring their systems are secure.. “[Y]ou’ve got all of these government departments heavily reliant on their service providers,” says Howard. Without proactive testing of the service providers, and monitoring the government systems for breaches, the only way the government will know if there has been a breach will be when someone from the wider intelligence community “comes and taps them on the shoulder saying ‘hey, potentially you’ve got an issue here.’”

Asked how hard it would be for New Zealand’s security services to track, in real time, the activity of foreign actors in our digital spaces, Howard says that would depend on the nature of the digital space. Setting up proactive tests to monitor for interference in the government’s own networks is one thing, but tracking foreign efforts to manipulate discussions by New Zealanders of public policy issues on Twitter, for example, would be more challenging.

“How do you know which ones are foreign? If they haven’t set their location to New Zealand (which can be arbitrarily set anyway), and they’re engaging in conversation, you don’t know where in the world they are. There are ways to determine this. The ways they did it on Russian influence in Brexit, they looked at the time of day that the posts were going up, and it was Russian business hours. They looked at a particular account that had sent 115,000 tweets, and all of them were in Russian business hours. They also looked at that particular user, and at the pattern of things they’d been talking about, over the course of three years. Brexit here, and then Trump there, Anti-Vax here, whatever. Whatever agenda that they’d been pushing at the time.”

So it’s possible to gather what could be called circumstantial evidence pointing to an account being based outside of New Zealand, or being used for political manipulation. But, as Howard says, it can be really hard to be certain. You can attribute on a micro-level, he says, if you can find tweets about local weather or politics, or something else that might give away who they are, and where they are. “But otherwise if you’re looking at the ISP layer, all you can see is, ‘is the person on Twitter, yes or no?’ You can’t see what the account is, what the username is, because that’s all encrypted. So all you know is whether or not this person at this internet connection, at this address, was using Twitter at that time.” Internet Service Providers don’t have access to the detailed information about what was done on twitter, says Howard. If for example they wanted to see who had viewed a particular video however (not on Twitter) . “They can run a query across their network and say ‘these are the people that viewed this video on this website,’ assuming the website was unencrypted.” Typically they will only have information to the level of “someone visited this website” however, they won’t know what content was viewed.

The heads of New Zealand’s security and intelligence agencies have publicly confirmed their assessment that foreign interference in New Zealand politics and elections are possible. At a public briefing in April, NZ Security Intelligence Service director-general Rebecca Kitteridge reported that her agency was aware of “efforts to mask the origin of political donations, covertly influence MPs and exert pressure on expatriate communities.” At the same briefing, Andrew Hampton of the Government Communications Security Bureau raised alarm bells about any plans to trial online voting, referring to international

evidence “that state actors have scanned electoral systems looking for vulnerabilities, so I wouldn’t say it’s given me reassurance, it’s probably given me more reason for concern.”

The government says it is aware of the risks of digital interference in elections by hostile states, and concerned to find ways to both identify and respond to any such interference as quickly as possible. One of the challenges, according to government officials, is that the process generally followed in order to confirm attribution of digital action is slow. By the time a post or action has been attributed with confidence, it may already have done harm. This risk is heightened during election campaigns, as was seen with the disinformation campaigns targeting Hillary Clinton in the days and weeks before the UN presidential election in 2016. By the time those actions had been attributed to hostile foreign actors, they had already done irreparable damage to Clinton’s campaign. Government officials say that finding a way to attribute quickly enough to prevent that kind of harm is a high priority for the government’s cybersecurity advisors.

Simon Howards says the solution to this challenge may be as simple as asking whether you need attribution. “Do you need to attribute? In [the case of Hillary Clinton] where there was an information dump, there wasn’t very much in it that was bad for her anyway. It was just the fact that Trump jumped on it and said ‘Crooked Hillary and her emails.’” So part of the problem with interference, says Howard, is that in the end politics isn’t even about facts. Even if you could attribute it, once the story has momentum, the facts won’t slow it down. “If it’s a long-term campaign of influence, using fake news, for example. You don’t want to repeat it in order to debunk it or attribute it to a bad faith actor.” This kind of situation needs a different kind of solution, says Howard, one that doesn’t rely solely on facts to slow down a manipulative use of disinformation once it has started to spread, and “you would have to have media on board. It would be quite difficult to accomplish.”

And, finally, as Marama Fox pointed out, the risk of false information being digitally published and spread in the heat of an election could have a devastating effect on a candidate even if there was no suggestion of involvement by foreign state actors. “I remember one person who rang up to someone on our campaign and said ‘I’ve just been with Marama all night at the casino, the cops were after us, we just spent \$10,000 of the Māori Party money, and we’ve blown it on the tables.’ She was just jack talk, but if she’d gone to my Facebook feed, and said that, in the two days leading up to the election. Anybody could do that, at any time, and that’s dangerous.”

MANIPULATION THROUGH POLITICAL ADVERTISING

Setting aside foreign actors, much has been made in the international media of the potential for political parties themselves, and third parties hired by political parties, to use digital media to manipulate potential voters. One of the main concerns is the potential to micro-target political messaging to people based on the personal data that the platforms have collected about them. It would be possible to find and target people whose previous activity on the platforms might reveal a susceptibility to, for example, extremist political views.

Joel Colón-Ríos says this potential for manipulation by political parties is why the mass accumulation of data by digital platforms is not only a matter of individual privacy, but also “a political, or a community concern.” The data that platforms hold about someone might seem relatively apolitical, like the fact that someone likes recreational fishing. But that can be used for political targeting, says Colón-Ríos. “They can target that person with

something that says x party has an attractive policy about recreative fishing and then follow that up with more general information in support of that party.”

Most interviewees for this project said that although the tools for this kind of targeted political influence are available to political actors in New Zealand, they haven’t seen much in the way of their use. While the potential for manipulation exists, says David Farrar, “overall, most of the microtargeting I’ve seen political parties and others do is really sensible.”

Anna Dean disagrees, pointing to the National Party’s campaign of ‘Let’s Undo This’ which targeted the current government. She says the campaign was “all about collecting people who are that upset about what is happening. It’s incredible, it looks exactly like the Labour Party branding.” Her view is that the use of ‘fear mongering’ in this campaign was incredibly effective. The nature of the advertising material used, together with the use of micro-targeting to identify people who might be inclined to be persuaded by these ads, has the potential to be extremely effective as a tool of political persuasion. Even if not widely used in New Zealand, Anna Dean argues, it is important these attack strategies are called out and discussed in the New Zealand context.

Andrew Burns points to Twitter to argue that there are some bad faith political actors online. “I suspect it’s just a very small number of dedicated political operatives who are doing it. Recently on Twitter a woman tweeted me about something political. The top of their bio was ‘concerned Hutt resident,’ which sounds like a parody anyway, but someone else tweeted a reverse Google search of their profile pic and it was a stock image.” The account was engaging quite ‘dramatically’ in New Zealand politics on Twitter, says Burns, making provocative comments and trying to engage people in arguments.

If this account was being operated by someone whose motivation was enjoyment of the disruption they caused, it would be called a troll account. If the account was not operated by a human, but automated by software, it would be a bot. But if the account is operated by a human who is working on behalf of someone or something else, with the motivation of disrupting political conversations, pulling political actors into bad faith debates or provoking extremist views, it would be called a shill. The term shill is someone who promotes something or someone for pay without divulging that they are associated with the entity they shill for, inherent in the definition is the idea that the shill account is not only hiding their identity, but also the party in whose interest they are acting.

It’s not always easy, or even possible, to distinguish between a troll and a shill account. They often behave in very similar ways, albeit for different reasons, and their impact on democratic conversations may be very similar. This is particularly true when high profile people who hold extreme views rally their followers to attack people on Twitter who have displeased them, as Stephen Molyneux did when Auckland Mayor Phil Goff denied him access to the Town Hall in 2018. The Twitter users who respond to these calls may not be shills, they may be acting on their own behalf and motivated by their own views, but the ability of people like Molyneux to mobilise them in semi-orchestrated ways means they can have a similar impact to, for example, an organised group of shill accounts acting on the instructions of a political party.

Nat Dudley agrees with other interviewees that we have not seen a lot of that kind of organised political manipulation of shill accounts in New Zealand. “But one of the things that makes me worried is New Zealand is pretty vulnerable. The first time a political party decides to dedicate a significant amount of resource to this, it’ll completely change the way it works.”

Marcin Betkier thinks New Zealand might be particularly vulnerable to this kind of operation because trust is higher here. “New Zealand is more trustful, more unaware. [I]t’s not really naivety, but New Zealand is still an island community, so we are stronger on personal relationships. [But] we have to be prepared for what’s outside, and in the context of democracy, for example, there is no barrier to election manipulation. Because we are on the same Facebook, on the same Google, and the rules are the same. [I]t’s just a question of who will pay for it, or whether there is anyone interested. The tools are already there.”

Andrew Butler also raises questions about whether our current regime for regulating electoral spending and advertising is adequate for tracking the ways in which political parties use third parties to advertise on social media. If it’s not adequately regulated, he says, and some parties are taking advantage of that, the incentive then exists for other parties to do the same. “If it’s not regulated and your opposition is getting an advantage by operating in a way that you don’t think is ethical or right or proper, and the outcome of it will be on a short term basis that you lose, what are you going to do? Of course, there will be a strong temptation to do the same as them notwithstanding your reservations.”

The New Zealand Electoral Commission regulations require that if a third party promoter of a political party spends more than \$12,000 on a campaign, they must register with the Electoral Commission. In addition, third party promoters may only spend up to \$315,000. Whilst there are clear rules around the use of social media, including on election day, the Electoral Commission does not track the activities and spending of third parties on social media, they investigate complaints if they are raised, meaning there could be a lack of oversight over third party social media spending.

The flipside of the potential to use digital platforms for political manipulation, if you have enough money to do so, is the way in which it is becoming harder and harder to use those platforms to reach an organic audience of people who have opted in to see your content.

Nat Dudley reported on conversations she’s had with people responsible for digital media for mainstream media outlets who are finding it harder to get organic reach for their content. “[Y]ou don’t get to choose what stories are popular on Facebook and what they put in front of people’s eyes. That’s just not how it works.” If news outlets do choose to boost a particular story, which isn’t already getting engagement or reach through the algorithms, that has a negative impact on any other promoted content from their page. The algorithms are tuned to give preference to content that gets attention. So if a news outlet spends money trying to put a serious piece of public interest journalism in front of a bigger audience, despite it not getting as much organic attention as, for example, some funny, alarming or shocking content - all their other promoted content will get less reach, “So you’re penalised for trying to promote things people are not engaging with.”

ONLINE ABUSE

One of the reasons this research was scoped so widely was to find out whether the major opportunities and threats posed by digital democracy were different in New Zealand to elsewhere. Global commentators on the issue had placed a lot of emphasis on the risks presented by fake news and foreign interference, while online abuse and harassment had attracted less attention, at least until the Christchurch mosque attacks. One of the hypotheses of this research was that, in the New Zealand context, online abuse and harassment might be having comparatively more impact on democracy than was generally recognised in the global commentary. Not because we had more online abuse in New Zealand than elsewhere, but because intentional disinformation and foreign interference, while both posing very real and substantial potential risks to New Zealand's democracy, did not yet play the role in New Zealand politics that they did in other settings, like the United States, the United Kingdom, India and Brazil.

This hypothesis was largely borne out in the view of interviewees, but as has been seen above, many warned that there was nothing standing in the way of New Zealand becoming more of a target for political manipulation either through intentional disinformation, foreign interference or a combination of both. Most interviewees were clear, however, that online hatred, abuse and harassment were already significant problems in New Zealand, and could see risks to democracy as a result.

One of the anti-democratic features of online abuse, notes Sandra Dickson, is that the people predominantly targeted are people who have already been marginalised in public discourse and politics. "[T]hey're the people of colour, they're women, they're queer people, they're trans people. And the level of vitriol and hatred that's directed at people is incredibly harmful."

One group who Dickson sees being targeted in particular are women journalists, particularly those who are covering "stories that have relevance to power." Attacks on journalists generally have democratic implications. Online attacks targeting journalists from previously marginalised groups carry the additional risk of silencing voices which have only relatively recently been given space in mainstream.

If you live at the intersection of being - for example - a woman in power and a person from a minority or marginalised group, Dickson says, you are likely to get even more abuse. "Golriz Ghahraman for example, the stuff she gets is just disgusting. And the level and volume of commentary about Jacinda Ardern's mothering, and whether that's impacting her ability to be a good politician, that's been amplified by digital platforms."

Ghahraman herself has seen a link between misinformation, disinformation and online abuse. One form of misinformation which was published online about Ghahraman was allegations disputing her status as a former refugee. Based on research about combating disinformation, she saw there was little to gain in directly correcting the misinformation. But she notes that the way in which the misinformation targeted her, focusing on her ethnicity and background as a refugee, is a sign of the overlap between misinformation and harmful communication.. "[I]t was the meeting point of mistruth, undermining credibility, creating doubt."

Swarbrick says that two of the features of digital communications that make online abuse particularly vitriolic and harmful are, firstly the dehumanising that the distance of digital media allows. "[W]hen you have some level of profile, for lack of a better word, there is this presumption that you're a concept, not a person. Because people are talking about

you like you can't read it." The other factor that makes online abuse so harmful, she says, is its integration with every aspect of your daily life. "[I]t's on my phone, and I can read it anywhere, 24-hours a day. It's stuff that people wouldn't say about somebody who they disliked, or who they knew personally in a public space, but because you're a public figure, you're a concept."

This kind of online abuse has a real impact on the people who are targeted. It can undermine their sense of personal security. Both Ghahraman and Marama Fox say there have been times when the level of abuse directed at them online has led to them needing to seek the support of Parliamentary security services. In Fox's case Parliament Security advised her team to call the police to ensure she got home safely some nights after particularly vitriolic and personal abuse had been directed at her online.

Some interviewees gave examples of women online who faced significant abuse and handled it particularly skillfully. David Farrar pointed to Chelsea Clinton, who responds to the most vitriolic comments with grace. "[S]he must have a filter for the most hateful comments, and she meets them with charm, and kindness. I'm sure it's hard to do, but it does actually have an impact." Other interviewees pointed out that while Clinton's approach is impressive, it's neither realistic or fair to place the burden of countering this kind of abuse on the people being targeted.

Erika Pearson argues, for example, that the platforms themselves ought to take a more proactive approach to countering this kind of abuse, but in a way that acknowledges and respects the experience and perspective of the people who are being targeted. This is particularly important, Karaitiana Taiuru has argued, when platforms or regulators are dealing with Māori victims of online abuse.

LACK OF ACCOUNTABILITY AND ANONYMITY

When you start talking about the risks and challenges associated with digital and social media, someone will almost certainly suggest that anonymity is the problem. In discussion of online abuse and harassment, and the ways in which those behaviours are making it unsafe for people to engage in conversation in digital spaces, one common response is that you shouldn't be able to comment anonymously. But interviewees were clear that anonymity needs to be approached with care. Democratic cultures and mainstream media have long provided the opportunity for some people to participate in public conversations anonymously. So the question becomes, how can we mitigate against the worst effects of that while protecting the right to be anonymous?

Author Nicky Hager sees anonymity as one of the driving forces of the shift towards more aggressive, unrestrained abuse and harassment online. As a general rule, he says, most people should be identifiable, because "they're just going to be infinitely more reasonable. They're going to be different. They just have to, they wouldn't keep their jobs if they wrote the kind of stuff they write. They'd get in trouble for it. It would reflect on them."

Hager's view is that there should be a shift to non-anonymity, but he's found that when he talks about this he comes up against an argument from young people, in particular, that loss of anonymity would mean they wouldn't participate in political debates online at all. "Why? Because they are scared of being seen participating in politics and having

views. Which is a terrible situation. Imagine if most people thought that they have to be anonymous to have a say in politics. It's really creepy. It's defeat. It's terrible."

'Real name policies', which is what Nat Dudley calls the idea that everyone should be required to use their real name online, often privilege people whose identities are socially acceptable. She explains, "that generally means that they're part of majority groups. They're gender aligned with what society views their gender alignment is, their sexuality is the more socially acceptable sexuality, all of those types of things."

Unless they are malicious actors, Dudley says, the reason people often want to be anonymous is that they are protecting themselves. They want to be able to talk about something without having people know and associate it with their identity. A really good example, says Dudley, is gender. Some people may not be openly out as non binary gender, and may not want people to be able to identify them especially if they work in conservative industries. But at the same time the online space can be very important for people to be able to talk to other people who share similar experiences.

"Being able to talk to people who share those experiences without having to have it public to the extent where employers or the media find that information about you. If you force people to use their real name for any discussion to have online, you deprive those people access to support networks that they need. And that can in turn make them the target of the malicious actors you're trying to protect them against."

Gwynn Compton agrees that there is definitely a good reason for people to have anonymity online. "Because there are some people who have valued views, who won't be able to share them because of where they work, or because they might become a target of hate." But the challenge is making a distinction between people like that, and other people - those who Dudley referred to as malicious actors - who might pretend to have legitimate reasons for anonymity. "People who sit behind two to three dozen different accounts and create their own fake crowds. It's a really hard one."

We've always had a place in society for authors publishing under pseudonyms, Compton argues. "There's always been a role for it. The difference now, is that previously you had to have the newspaper publisher, or the broadcaster, or the editor who knew who it was. And they could make the call of whether there was a legitimate reason for anonymity. With social media, we don't have that editorial role. So the question is who plays that role? Is it Facebook and Twitter?"

Andrew Geddis is a Professor of Law and Politics at Otago University. Like Compton he observes that while there has always been anonymity in publishing, there was always at least one journalist who knew who their source was and protected their source, or an editor who knew who had written the anonymous column. "So, there were, and are, very tight and very deeply held principles around not revealing that."

John Edwards also emphasised the need to be able to protect anonymity. "In many communities the right to anonymity is almost a precondition to engaging, to accessing a number of other rights. It's probably a precondition for all kinds of political speech, so the ability to engage anonymously is an important part of political and civic engagement."

But, he points out, the right to anonymity has limitations and those limitations vary in different contexts. In New Zealand we might have a higher tolerance for requiring people

to declare themselves because, Edwards says, we imagine ourselves to be a tolerant, safe place. But people travel. A story that illustrates the risks of exposing the identity of people online was the Ashley Maddison hack, after which there were blog posts by gay men from Saudi Arabia who had used the site as a means of hooking up when they were in the US concerned that if their identity was exposed back home, they could be executed.

As Edwards points out, at the root of that story is a society that penalises people for their sexuality, but protection of anonymity and privacy online plays a critical role in keeping people safe. “But privacy can’t be a shield by which you infringe upon other people’s rights by hounding them to suicide.” So there is a delicate and critical balance to be struck.

Even where they are not anonymous, says economist Shamubeel Eaqub, people’s behaviour online is different to offline. “They might not have actual anonymity but they are still nastier with what they say on Twitter compared to real life, because they feel less accountable. I find [Twitter to be] a very angry, snarky and a negative space. They’re all really angry and snarky. And then you meet them in real life, and they’re really lovely people. Social media, says Shamubeel, allows people to operate in a different way, without the social constraints we’d normally operate by. “So the rules of the game are quite different online, than if you and I were sitting in a cafe talking about something.”

SILENCING

One of the harmful impacts of online abuse on democracy, says Golriz Ghahraman, is the way it can silence already marginalised groups, excluding them from conversations on matters of public interest. “It puts some people off participating in democratic processes. It puts some people off participating in the conversations online about issues of public interest, whether it’s domestic violence or whether it’s foreign affairs. Whatever it is, some people are just going to say ‘Well I’m not going to participate because I’m a person of colour, or because I’m from a migrant background’.” Ghahraman says when she raises her concerns about this with people, they often have “very disappointing aspirations in terms of what we could do about it, and it seems to be that people have given up on the idea that we could regulate these spaces at all.”

But, she argues, the government has a responsibility to be proactive in tackling these issues and ensuring that if digital spaces are going to be used as an integral part of our democracy, that they are safe for everyone. If not, she says “there is a very strong case for people to opt out of our democracy, and quite frankly that’s the intent [of the abuse].”

Thomas Beagle also sees particular threats from online abuse to participation in public debate by people from marginalised groups. Trolling has become, says Beagle, a tactic for driving people off the internet. “That’s a concern to me because being able to drive people off the net is a real threat to our freedom of expression. The net is where we’re having our political debate.” And the people being driven away by online abuse, he says, are the people who are already disadvantaged in some way. “Of course in our society, that means Māori voices, LGBT voices we’re pushing off. Women as well, some of the misogyny we’re seeing here is a real worry.”

As one participant put it, there is a risk of a ‘multiplier effect’ in terms of equity of democratic access. “Already structurally, minority and vulnerable groups voices are not proportionally represented in government, or the media. I may be disenfranchised in terms of my ability to voice my opinion offline already. So now perhaps it’s more important that I do that online. But I get attacked for what I look like or something. So there’s a kind of multiplier effect from that, in terms of my equity of access.”

Sandra Dickson also sees that online abuse of women in politics could have the effect of putting other women off getting into politics, “I think it absolutely has an impact.” Although she is quick to point out that the positive impact of having a female Prime Minister probably far outweighs the harmful impact of seeing the online abuse directed towards her.

One of the themes that emerged in the interviews about online abuse was the pressure in New Zealand to be stoic about the effect of the abuse. As Dickson put it, “we have the cultural norms in New Zealand around just getting on with it, ‘she’ll be right, mate’.”

Swarbrick agrees. If she were to complain about all the abuse she gets, she says, people would tell her ‘if you can’t stand the heat get out of the kitchen’. So you have to prove yourself. “You have to consistently know more, have done more research, and be impeccable, because if you stumble, you stumble for all of the people who you supposedly represent.”

DISENGAGEMENT AND ATTENTION HIJACKING

PLATFORM DESIGN

Tech reporter Richard MacManus observed that design features of the platforms influence the nature of engagement on those platforms in ways that undermine democratic discourse. Twitter, for example, works best for “black and white opinions,” initially because of the character restriction. “You have to kind of truncate your arguments into sound bytes and each sound byte has to live on its own. That leans towards having to be very specific, and black and white in what you want to say in each tweet.”

Historian Paul Moon noted a similar pattern. Because you have such a small space in which to make a statement, especially on Twitter, “the tendency is to have something that’s fairly short, compact, pithy, and packs a punch to make the point.” This leads, according to Moon, to more rhetorical statements because “the shorter the statement, the more rhetoric rich it has to become to compensate for lack of substance.” So if someone is interested in politics, but getting most of their political information, news and commentary from blogs, posts on Facebook and Twitter, then you will be set up to have a view of politics that is “reduced to these easily digestible bits”, says Moon.

But the design of digital media doesn’t only influence the substance of the content posted through word limits. It also does so by rewarding people for engagement. Because their business model runs on the ability to attract and retain people’s attention, digital platform companies employ teams of psychologists who design, test and iterate features that provide almost irresistible physiological and emotional rewards for getting people to engage with your content.

So, as Moon says, the design of these platforms don't only simplify content, "it also simplifies it for the purpose of a reaction. In a sense the argument is accompanied by the need to get someone to react to it. So the test of its importance is partly the extent to which people react to it, and maybe agree to it."

Erika Pearson has also looked at the impact of digital media design on the language and etiquette used by people on the platforms. Each platform has its own norms, although people from different cultural backgrounds will use them in different ways. "Twitter is a real rage chamber and people accept that you can rage today and share a recipe tomorrow and you are not held to account for yesterday's ramblings the same way as you would be in a news article or a print article or a Facebook post perhaps."

Pearson points out that, although rage has always existed, the normalisation of raging on social media is quite different to traditional and print media or offline public spaces. "The expectation is that online comments are where you vent and there is an acceptance of that and you are not held to account the same way that you would be at a town hall for example."

Pearson herself said that she shares very few opinions on Facebook, and that it would take a lot to get her involved in a conversation because "you do get keyboard warriors and you do get people who are less articulate and versed in actually crafting argumentation in a reasoned way." On the other hand, she reflects, you can't yell over the top of each other the way you can in a face-to-face situation, "so maybe it is levelling in that sense. You can jump in a conversation where your voice may not be heard – quite literally heard – in a more face-to-face environment."

Low levels of active participation in digital spaces, says Anna Rawhiti-Connell, are normal. In digital engagement practice, she says, there is a rule of 1:9:90 which refers to the ratios of engagement in most online spaces. "So 90% of people who dip in and out of those spaces will just observe. Then 9% of people will engage in some sort of light participatory actions, but it's actually 1% of people who drive the bulk of the engagement. It's 1% of people who become super-users, the 1% of people who will try and respond to everybody. And it's that 1% of people who will complain, or advocate. So you've got 99% of people just passively engaging." All of which challenges the idea that social and digital media are creating some sort of new space for a more participatory democracy.

THE ATTENTION ECONOMY

At the foundation of many of the design features of digital media is a business model that trades in attention. Some have called this the 'attention economy'. The core business incentive of the big digital media platforms is to get - and keep - your attention and then sell it to advertisers.

As Ian Apperley put it, "all that any large internet company wants, whether it's Facebook or Google, is to be the front page of the internet. So it's the first thing that you do when you wake up and open your eyes in the morning, they want your eyeballs. They'll do anything that they can in order to get there."

In order to achieve this, these companies have hired perhaps as many psychologists as they have developers. Which is arguably that's why their tools work so well, because they understand human psychology and are willing to invest in developing and testing

applications that work with the idiosyncrasies of the human brain, in order to maximise attention. As Apperley puts it, “it’s just basic human psychology. This is the way that people communicate. Three fundamental, backbrain principles: Staying connected, are you part of a tribe? Breeding. And survival. That’s it, at a really fundamental level.”

In the process of maximising the vulnerabilities of the human mind to secure our attention, the platforms have - inadvertently - built a system that is perfectly calibrated to accelerate the spread of content that alarms and outrages us. Unsurprisingly, given that these powerful platforms dominate the distribution of information and public conversation, some commentators have labelled this the ‘age of outrage’.

As Gwynn Compton reflected, the design of the platforms - intentionally or not - creates incentives for campaigners and communication professionals to use provocative content to draw people into engagement with their brands or pages. They learn to work with the algorithms in order to get their content in front of the audience they are trying to reach, and that means using outrage for organic reach.

Compton recalled two examples from when he was working for John Key. “The first was shortly after Andrew Little had become Labour leader and he told John Key to ‘cut the crap’ over something, I can’t remember what, which Labour ran with and it got a great reaction online. Later on, when ISIS was on the rise in Iraq, John Key took a similar approach when he told Andrew Little to ‘get some guts’ about the need to send NZ troops to Iraq. Both Labour and National used these provocative comments because they were fiery and confrontational and they would get reactions out of those following our respective pages.”

Part of the reason to run with confrontational content like that, Compton explains, is because if you get great engagement on one post, that brings to the fore your other, more serious content about policy, for example. It’s a way to work with the algorithms to get greater organic reach for the kind of more serious content which is unlikely to generate a lot of engagement on its own. It might seem like a terribly cynical thing to do, reflects Compton, “but you’re operating in an environment where you need to get people to see your content.” And in New Zealand politics, he explains, if there’s not enough money to pay to advertise at a level that will cut through the noise, that then that’s the way to do it. “You get people to engage with something that’s provocative.”

Economist Shamubeel Eaqub also sees significant risks to a democratic culture in the way digital media has been designed to maximise attention by engaging the “more primal part of the mind.” Digital media is designed to deliver “the endorphin hit, the adrenaline hit,” says Shamubeel, and that comes from “the quick reaction, the witty repartee.” You don’t get that from nuanced conversation, “it’s like you have to be polarised.” This is a problem for democracy, says Eaqub, because democracy needs to be able to solve big and complex problems. “And inevitably big and complex problems have big and complex solutions, quite often spanning lots of grey spaces, and that requires us to make compromises, and give some things up to do other things that we think is right.” That is much harder in the online space, says Eaqub, where the need to differentiate yourself can lead to “quite polarised and quite aggressive” behaviour.

None of this happens by accident, say James Every-Palmer. “Facebook has carried out more randomised control trials than in the history of medicine because it’s so easy. They just can do hundreds a week where they give one set of users a particular experience and

another set of users another experience, and what they're fundamentally measuring is how sticky, how addictive that makes the site." All of this is designed to get people to spend more time on the site. It's designed to get and hold your attention, which can then be sold for advertising revenue. The business model, says Every-Palmer, is psychological addiction. It's not designed with the benevolence of being democratically friendly."

Alongside the addictive effects built into the design of platforms like Facebook, Ganesh Nana observed that the design of digital media leads to a level of information overload that makes it very difficult for people to be critical consumers of the information they receive. "[Y]ou could argue that consumers are not that discerning because of the swamp of information that they have basically thrown at them, so not only can they not distinguish between good and bad information, they are not actually interested in distinguishing between them because they just have so much." This undermines a key assumption about how markets function, says Nana, because if consumers can't distinguish good information from bad, "they can't pay for the good stuff and therefore not pay for the bad stuff."

Erika Pearson agrees that people opting out of digital media are more likely to be doing so in response to overwhelm than as an act of apathy.

Elyse Robêrt has observed that people of her generation, in their late 20s and 30s, are decreasing their engagement with social media platforms, with many leaving, partly as a result of the noise, clutter and information overwhelm of the platforms, partly as a result of frustration with changes to the design of the platforms, and partly in response to abusive or aggressive behaviour on the platforms. "We are dropping off. We are not using it like we used to. We are fatigued, we are exhausted, we are just irritated." She says it would be interesting to know where those users are going. "If they are getting pissed off and fatigued with the social media they helped build, what are they craving and where do they have those conversations?" She thinks a lot of people are moving to closed groups, to avoid 'flack'. She explains that she personally would very rarely share anything political on Facebook or any other social media."

DISTRACTION AS A THREAT TO DEMOCRACY

Nicky Hager identifies the distraction of digital media as one of the greatest threats to democracy. The ability of citizens, and for that matter journalists, politicians and officials, to think clearly about complex issues is critical to a healthy democracy. Digital media, says Hager, is undermining that ability. "I watch other people walk out of a meeting, get out of bed even, and immediately start to look at what's happened on their phone. While I'm with other people, their phones are ringing. Texts are coming in." It is a constant source of distraction, and as Richard MacManus says, is creating "a kind of a subtle shift in people's ability to do long term thinking."

What makes this even more difficult, says Hager, is that people can't easily leave digital platforms because of the role they have come to play in social connection and inclusion. This is especially critical for young people. "When we're teenagers, not fitting in with our friends is truly terrible, and harmful. So if every party, every joke and reference, if everything is being communicated by Facebook, and other platforms. It's a really hard choice. I'd say it's a virtually impossible choice."

Ganesh Nana builds on this by pointing out that the platforms have been designed to be addictive, and "once you get into the addiction route, there's no way your decisions are rational."

LOSS OF PRIVACY

Another common concern about the impact of digital media on democracy is the way it has eroded personal privacy. Privacy is crucial to democracy, says John Edwards, “because the secret ballot is a fundamental part of our democratic system, and democratic process. People should have the right to have freedom to hold their own beliefs and not be required to share them with others.”

As increasing amounts of human communication, and activity, are digitised this increases the state’s ability to monitor citizens activities, including activists and social movements, to censor speech and to track communications. If everything political activists say or do can be intercepted and monitored, this is likely to have a chilling effect. This is a concern for all citizens, with particular risks for human rights defenders and journalists - given their particular roles in holding governments to account. Protecting the right to privacy for all citizens is therefore crucial to a healthy, functioning democracy.

The Special Rapporteur on the Right to Privacy, Joseph Cannataci, says that privacy is one of the three essential rights necessary for the proper formulation of personality. He argues that privacy is both as a right in itself, and also an enabler of other rights. In addition to enabling freedom of expression, he asserts that respect for privacy is essential to the unhindered and free development of one’s personality. In other words, the ability to keep some information private or limited to certain groups is said to be fundamental to individuals’ ability to develop a distinctive character, sexual identity, life path, aspirations. All of which are critical to the development of the ‘alert, informed citizens who are capable of making important moral judgements’ who are needed for a functioning democracy. As John Edwards puts it, “without a private space to think about and assess and analyse the things that are happening in the world, it’s more difficult to participate in democracy.”

One simple take on privacy is that consumers need to proactively and consciously manage their own privacy. As Member of Parliament Melissa Lee says, “everybody should be concerned about how their data is being used and what kind of permission they are giving when they sign up to an app or media platform. The fine print is really important.” Having said that, Lee concedes that “most of us don’t read the fine print. That isn’t a license for big media giants to abuse the trust.”

Human rights lawyer John Hancock, expands on this, saying that the legal consent process relies on the person who is consenting being able to make an informed decision. In the case of many situations in which people are asked to give consent for data to be gathered by digital media platforms and applications, however, he says “there’s no real process of engagement that gives you a chance to really think about where that information is going.” In fact, he says, a lot of that personal data will be stored outside of New Zealand, so that if you wanted to make a claim to it, if it was tampered with, “it might be difficult to be able to use your own laws to be able to respond to it.”

Kathleen Kuehn is a Senior Lecturer in the School of English, Film, Theatre and Media Studies at Victoria University, and is the author of *The Post-Snowden Era: Mass Surveillance and Privacy in New Zealand* (2016). She says that current data privacy laws are based on the assumption that people understand and make a rational choice about sharing their personal information, but this is the reality in the digital environment. “To expect this level of ‘informational self-determination’ is unrealistic in a surveillance society,” says Kuehn, “Rather than seeing contemporary surveillance as a series of individual privacy invasions then we have to see them as part of a new landscape of social power.”

CONSENT FATIGUE

While a few interviewees argued that users of digital media platforms choose to give up their privacy in an informed exchange of personal data for convenience, more expressed concerns that the consent approach to data management was no longer working to protect the privacy of citizens, including privacy from the state.

Privacy Commissioner John Edwards argued that consent in the digital age had become degraded through repetition. “Every day people are being asked to click, click, click, to show that you consent. And that I think has degraded consent. I think people do get this kind of click fatigue. You simply can’t read all the statements.” This fatigue, he suggests is at least partly by design. “The incentive [for the companies] when you have that consent based culture, is almost to make it more complex, to make it too difficult to actually read through, and to just revert to the click. ‘Here’s a sweet reward, you want to play Candy Crush? Just click here!’ And then you get your reward instantly.” So the degradation of consent is built into the design because the business model depends on accumulating as much personal data as possible. And opting out of giving consent to digital platforms, says Erika Pearson, “often carries a social penalty.”

TRADING OFF PRIVACY FOR EFFICIENCY

Alistair Knott is Associate Professor of Computer Science at Otago University and is an expert in artificial intelligence. He says there is inevitably a trade off between the efficacy of the platforms and protection of privacy. Are we willing to reduce the convenience of these tools in order to strengthen the protection of our personal data?

For example, he cites the algorithms that drive Google searches, YouTube recommendations and Facebook feeds. “Anywhere where you are giving a feed, there is something ranking the items that can be presented in that feed, and the ones at the top are the ones that it thinks you are mostly likely to like or most likely to click on.” And there are feedback loops operating, whereby the user is most likely to click on one of the first items presented to them in a search or feed, which reinforces the algorithm and means they are more likely to be feed more content like it.

All of this is very convenient if I’m searching for a restaurant, or looking up movie times. Google’s algorithms predict that I’m probably going to want to see the website for the cinema I have clicked on or searched for in the past. The fact that the algorithm has stored data about my previous searches makes my search today more efficient.

But, as Knott points out, this efficiency becomes more problematic if I’m seeking information on a policy issue. And, he says, it’s not inevitable. The algorithms are designed to present me mostly with options that match my previous choices, options that I’m likely to prefer. But they also generally present me with some new options, and then records whether or not I choose them. These options fall outside of the preferences I have already indicated, so they give the algorithm an opportunity to learn new things about my preferences.

Knott points out that the proportion of predicted preferences and new options is a parameter that can easily be adjusted in the algorithm. It could be ‘turned up’ so that we are consistently presented with more new options and information, relative to the options and information we are presented because we are likely to like it. This is a simple design

choice. Turning up the dial on new ideas and options might be better for democracy. But it could also undermine the software's efficiency at giving us what we want. Which might in turn make us less likely to use it.

As Knott explains, "[t]he irony is that the reason why Google is so good is that it does give you what you want. You know, it knows for instance that I am doing a search and that I am probably interested in the name of this restaurant in Dunedin where I live, rather than elsewhere. So, you would be ... you would arguably be messing with the functionality of the thing by doing this."

And while we know "that sophisticated algorithms are tracking and shaping our movements," says Kathleen Kuehn, "we are encouraged to forget it is happening as we take pleasure in its perceived benefits." So some of the challenges to democracy are built into the very design of the software, and those designs are there because they support the underlying business models of the digital media companies. A business model in which, as Erika Pearson pointed out, "you are not a user, you are a commodity."

Karaitiana Taiuru, PhD candidate at Te Whare Wananga o Awanuiarangi and long time advocate for Māori online rights, has also advocated for greater user control over personal data. He argues that Māori data; 'data that is held by Māori, made by Māori or contains any Māori content or association' is a taonga under Article 2 of Te Tiriti o Waitangi, and should not be treated as a commodity product. The current approach to data taken by digital platform companies doesn't account for indigenous rights, nor does it allow Māori adequate control or sovereignty over their own personal data and what it is used for.

POLARISATION, RADICALISATION AND 'ECHO CHAMBERS'

One view that emerges from these interviews is that the features of digital media that threaten democracy are baked into their design, which makes them hard to redress through updates to, for example, content moderation or data privacy policies.

James Maclaurin is Associate Dean of Research at the University of Otago, with a particular interest in philosophy of science computing and information science, particularly artificial intelligence and its impact on humanity. One of the dangers he sees in the way digital media allows people to be targeted is that people think they are being informed about something, so they don't go looking further, but they are in fact only seeing a slice of information which has been targeted to them based on their particular demographics or interests. "So, I think that this election is mostly about one thing, because, that is what I hear. And the people I know are like me and so they see and hear the same things, and we think it's all about that. So one of the issues about targeting is that it makes the country more tribal in this issue-by-issue way."

Joel Colon-Rios says people may engage on digital platforms in ways that confirm their own views. "When people are interested in a particular political issue, to give a random example, the death penalty and go to the internet to engage with other people interested in that issue, they may tend to go to websites that people who have similar ideas to them, frequent. And what tends to happen, according to some research, is that people, instead

of engaging in a form of deliberation in which they are likely to change their views if confronted with good arguments, they become more, and more polarised. That is to say, their initial views are strengthened: the internet allows them to discuss issues with people that have very similar views to the ones they have.”

On the other hand, says Andrew Burns, it’s natural to want to “stay in my little silo sometimes, because it’s exhausting. Reading horrible things online all day is not something that people want to do, and they should be able to opt out of that.”

Using the example of British tabloids, which offer their readership a version of the world that confirmed their existing biases and beliefs, Andrew Butler notes that people were able to seek out information sources that conformed to their worldview before the rise of digital media. Digital media, however, has enabled a “more extreme version of that.” In the days before digital media became a dominant source of news, he says, people would read their preferred paper in the morning but would also watch the TV news in the evening. So in the morning they might have their existing views confirmed, but in the evening they would see a presentation of the news which was simultaneously being viewed by most other people in the country.

“So whether you liked it or not, you would hear the other side. You’d turn on the BBC or ITV, and you’d say that can’t be right, I read in The Sun... So you’ve got your view but you’re also hearing what the opposite view is.” Importantly, Butler notes, you would have been hearing that other side directly from the people who hold that perspective. A reporter might interview someone who takes a different view to you. So you are not hearing or reading that view once it has been filtered through the lens of, and critiqued by, people who agree with you. Which is what is often happening on social media.

Thomas Beagle makes a similar point, that in the past we had, to an extent, a shared narrative of what was being presented in the media. “That shared narrative was often biased,” he says, “or incorrect. Even so we could talk to each other about it, and we could talk on common terms, and we could argue for or against it, whatever we liked.” That kind of common conversation has been overtaken by the fragmentation of social media, which makes it hard to talk to people who are not exposed to the same narratives, the same slice of information that you have been. “It’s like we’re coming from different worlds,” says Beagle, “we don’t have that shared narrative to base our shared conversation on, it’s like we’re complete foreigners to each other.”

On the topic of radicalisation through online platforms, Nat Dudley says that the internet, and particularly the big media platforms, are not tools to help people to see truth, but tools that help people to further the opinions they already have. “The entire premise of a recommendation algorithm is give you more of the things that you like. Combined with click bait headlines, this means increased radicalisation in whatever direction you’re already going.” You can see the results of this, says Dudley, in how young men become more and more radicalised. If you watch a video on Youtube, the recommendation algorithm is designed to keep showing you things that are like the things that you have liked in the past, but more so.

Another way in which digital media enables radicalisation, says Paul Spoonley, is by amplifying views that would never have been provided through the mainstream media, views that are unfiltered. “That’s the Alex Jones phenomena in the States, where he can hold a view that a mass-shooting was faked, which just seems extraordinary. In the

past, in order for him to communicate one-to-many, he would have had to convince media organisations that this is a legitimate view. He doesn't have to do that anymore, the gatekeepers have gone."

Gwynn Compton says that we know these trends are dangerous because we can see in them echoes of what we saw in 1920's-30's Europe. "You can see similar stuff happening now with the rise of far-right parties." We've been fortunate for a long time in New Zealand, says Compton, to not have any sort of mainstream far-right presence. Although he does see that all the major parties have 'dabbled in it'. "Winston Peters has stoked it occasionally. Labour toyed with it with the Chinese surnames. National's done it with the iwi vs kiwi over Māori treaty stuff or more recently with the United Nations Migration Compact. We've always had dabbling in it but we've never had it organised in the same extent as say Australia, and you do wonder at what point does that tipping point come."

LACK OF TRANSPARENCY AND AUTONOMY

Woven through all of these risks are two fundamental challenges which interviewees came back to repeatedly. Firstly, we are living in deliberately engineered online environments created by virtual monopoly tech companies into which we have no input into and over which we have no control. Secondly, we can't even really see how these environments work, which makes it near impossible to challenge or change them.

Many interviewees described a lack of transparency as being built into the design of these digital platforms, including through the function of algorithms. Tech reporter Richard MacManus says opacity is a big problem with the algorithms of Twitter, Facebook and Google. "Nobody quite knows how they work. Facebook itself has admitted a number of times that it can change the algorithms so people see certain things, and not see other things. And as a user it's hard to tweak things so you can see more of other stuff, and you don't know what that stuff is."

Nat Dudley specialises in user experience and product design and has a particular interest in ethical software design and data use. One of her concerns about the impact of the major digital platforms on democracy is the way they are designed to deliver very tailored individual experiences to users. Which means that everyone's experience of the internet and online media is individual to them. This makes it hard for any of us to get a good picture of what is actually going on overall.

"Especially for those of us who grew up with a more open web," explains Dudley, "It's very hard to see that because we're used to an internet where everyone sees and experiences the same thing and where the internet is a big broad open space that you can explore at will. And the change to that has been subtle but pervasive."

The change has come from a raft of different design features of the major platforms. Google tailors your search results so that what you find when you search will be different to what someone else using the same search terms would find. YouTube algorithms are tailoring your content recommendations. And your entire experience on Facebook and Twitter are shaped by your previous interactions and the preferences you have expressed through those interactions. Each person's experience is unique to them. "My Twitter is not your Twitter is not someone else's Twitter," explains Dudley, "and the same applies for Facebook."

This personalisation makes it hard to see what is going on in a wider sense, “because you can’t see what other people see, you can only see the experience you have.” And people are not even necessarily aware of how customised their experience is, because the algorithms making those choices for them are invisible.

Dudley sees a number of problems in this which are a challenge to our democracy. “Like the fact for many people, especially people who came to the internet younger or later often from lower socio-economic groups, the internet is Facebook. Sometimes it’s YouTube. But Facebook is often the entirety of the internet. So if they say look it up on the internet, they actually mean look it up on Facebook.” So the internet and all the information online, is experienced by many users as mediated through a platform owned by a corporation and designed to maximise profits.

As Dudley points out, these platforms were never designed for political action, or for the distribution of public interest journalism. “The motivation for Facebook was to get as much of your data as possible, and then try to make money from it, in their case by selling ads. It was never designed for neutrality or education or for promoting fair and accurate viewpoints. It’s just not part of their fundamental system of design. And the fact we’re using it for those things are where all those issues are popping up.”

We’ve come to rely on these platforms as our primary portals to news and information. In many cases people have come to think of them as the internet. But, she points out, “the big media platforms are not tools that help you to see truth, but tools that help you to further the opinions you already have. The entire premise of a recommendation algorithm is give you more of the things that you like. And that combined with the click baity headline type writing means increased radicalisation in whatever direction you’re already going. If you hold liberal viewpoints, it’ll recommend more liberal and more open, if you even start questioning the other way, it’s a real fast slide into some really scary stuff.”

The tendency to offer up increasingly extreme content is baked into the design of these platforms. It is fundamental to their business models, not because they are interested in radicalising their users, but because attention is their commodity. They trade in our attention, and they have trained their algorithms to do whatever it takes to get more of it. The fact that the human brain has evolved to give attention to content which is alarming is not the responsibility of the digital platforms. But they are responsible for having exploited our cognitive biases without regard for the impact that would have on our democracy. Particularly, as Dudley says, when that exploitation is combined with “the fact they’ve undermined media full stop, so that there are very few places people can go where they can trust they’re getting fair and impartial information anymore.”

One of the implications of operating in his engineered environments in which algorithms are deciding which information we see, says Marcin Betkier, is that we no longer have random encounters with information the way we can and do in our offline lives. For example, he says, if he chooses to walk home from work “maybe I’ll see something, I’ll meet some people or not.” This random factor doesn’t happen online. “People just go through this architecture of the online world, to parts which are somehow pre-designed, and they lose this exposure to something which is just random, just different. So they lose the opportunity to make changes in their lives, because of something unexpected, something random.”

These deliberately created environments are being built by huge corporations in which the people using the platforms exercise no control, because they are not the customers. Advertisers are the customers, and the platforms are designed to meet their needs, not the needs of the people who use the platforms to communicate with their friends or to read the news. “We don’t really know how our data is being used to create these environments,” says Betkier. “and we don’t realise how these environments impact us. But we do know that these environments aren’t being created in our best interests; they’re being created in the best interests of platforms and advertisers.”

A lack of transparency, says Nat Dudley, is always going to be the biggest problem with these platforms. “They’re not regulated in any way or required in any way to be transparent about what they do and how they do it. And they hold all the power in this equation.” Which is, she emphasises, a ‘really uncomfortable thing’. These companies hold the power over how governments interact with their citizens, how activists reach people, and “no one can see why they make the decisions they make or how those decisions work. It means that there’s a massive black hole in our democracy which we can’t see out of anymore.”

James Maclaurin questions whether transparency is even possible when it comes to the digital platforms. It depends what tools they are using, he says. “If it’s a relatively simple regression-based tool, then you would have those facts that you could use. If the tool is a deep learning tool, then at best I might be giving you some very approximate story about why you are getting this information.” Even if it were possible to make, for example, information about political advertising transparent to people, Maclaurin isn’t convinced it would help mitigate the impact of that advertising. “The literature in cognitive biases on general says that warning people of the dangers of bias has surprisingly little effect on their decision-making. It is just the way our brains are made. So, if I am targeting some feature of you that really is an ingrained part of your personality, telling you that you are being targeted might not do much.”

Colin Gavaghan is also skeptical about whether requirements on transparency in advertising will actually solve the problem of people believing false information. He says that in the context of fake news ‘we are all aware at a certain level that this is happening, but do we act like we are?’ He points to the example of people being aware that certain publications publish misleading or untrue content, but people don’t tend to remember where they have read something and tend instead to treat all sources of information as equally true. As a result, he wasn’t convinced that transparency measures or ‘labelling content as suspect would make a lot of difference to people because you don’t necessarily remember where you saw something.’

SOLUTIONS

Interviewees suggested a range of interventions and solutions to both maximise the opportunities for democracy presented by digital media and minimise the threats. These range from interventions at the structural and systemic level through to suggestions for individual behavioural change.

CHANGE STRUCTURES & SYSTEMS

At the structural level, most interviewees suggested regulatory and legislative changes, although most recognised that in order for regulation at the national level to have an impact on multinational tech companies, it would need to be embedded in an international system of internet governance.

MULTI-STAKEHOLDER GOVERNANCE

Many interviewees recognised that the success of any attempts to regulate digital media domestically would be dependent on some form of global cooperation. Some described a multilateral process similar to those applied in relation, for example, to international trade or labour law negotiations. Jordan Carter, of InternetNZ, however emphasised the importance of multi-stakeholder, as opposed to multilateral, engagement which would include participation by the technology corporations themselves, and by civil society.

As a principle, says Carter, internet governance is best done in a multi-stakeholder manner, not a multilateral one. “The idea of something like the United Nations launching into a treaty making process on content regulation, is a deeply terrifying one, because that then opens, at least in principle, the possibility that they’ll try to regulate the underlying technical infrastructure, which we’ve always tried to keep well away from governmental control.”

The UN Internet Governance Forum was established as a multi-stakeholder forum to discuss issues arising in relation to the governance of the internet. Carter says it hasn’t been as useful as some hoped at it’s foundation, but it at least is based on the principle of engaging a full range of stakeholders, including tech companies and civil society, in discussions.

For multi-stakeholder processes to produce results that would be in the interest of all stakeholders, they need not only to include a diverse range of stakeholders but also

to ensure that there is some balance of power between the different parties. One idea that came up in a number of interviews was the need to strengthen the civil society arm of multi-stakeholder internet governance, and in particular to need to better represent the needs and interests of users of digital media. Marcin Betkier and Erika Pearson both suggested that more needed to be done to redress a massive imbalance of power in these processes, and particularly to restore more power to the voices and interests of internet users.

This would also include ensuring that the cultural needs of Māori in online spaces are upheld, and that Māori sovereignty over the taonga of data is incorporated into new governance structures and regulations. Karaitiana Taiuru has identified two core Māori values which would support an inclusive data system; manaakitanga, where data users demonstrate tolerance and respect, and kaitiakitanga where New Zealanders, collectively, become guardians of our data ensuring it is managed with integrity in a way which enhances personal mana, rather than eroding it.

REGULATION

The arguments we heard about whether or not digital platforms should be subject to more regulation can be grouped around two core questions. Firstly, the normative question of whether regulation ought to be pursued and secondly, whether or not regulation in these digital environments would be effective, and therefore worth investing in.

On the normative question, James Every-Palmer sees a clear case for regulation. “[I]t seems to me that we have had blinkers on over the last decade. We’ve ended up with these huge social media platforms which are a significant part of our lives and can have lots of impact on our real world experiences. They give rise to a vast array of issues from fake news to objectionable content, to the collection and use of personal data, to addiction and mental health issues. And they are just simply unregulated.” There is no reason, he says, for them to be excluded from regulation.

Taking a very different position, historian Paul Moon says regulation isn’t needed. Digital media as it exists is an example of the market in operation. If there were a need for change, he argues, “the market would create that demand, but it hasn’t, so perhaps we need to go back one stage and ask why is it that people haven’t demanded it? I’m sure they’re aware of it, so why haven’t they? Perhaps because they don’t want to. There’s no appetite for it.”

Ganesh Nana responds to this argument by pointing out the many market failures that can be identified in the current model of digital media. These range from the impact of opaque design and overwhelm on users’ ability to make an informed decision about, for example, trading away their personal data for convenience, through to the anti-competitive effect of multinational platform companies operating in New Zealand without complying with the same laws and regulations that their domestic competitors have to abide by.

Dave Moskovitz makes the argument that regulation may not be needed as people learn and adapt to the new digital reality. “This whole digital thing is new for most people and it’s a fundamental shift in the way people think, and the way people interact, and we’re still learning. I’m confident that as a society we will learn good ways of respectfully interacting with each other, because otherwise the future is just too horrible to contemplate. But maybe that’s too head-in-the-clouds optimistic.”

One of the challenges of regulation, says Alistair Knott, is that the kinds of changes needed to mitigate against the threats posed to democracy by digital media are likely to make the platforms and products less effective at delivering the kind of convenience that makes them so appealing to users. As an example, he says that the algorithms that determine the content that gets delivered to users - whether as a result of a Google search or in a Facebook feed - are designed to give you more of what you have previously given signs of liking. This makes them very effective at delivering users with content that matches their preferences - a very useful quality if you are searching for a local restaurant. Less useful, in a democratic sense, if you are searching for information on a matter of public interest.

But there is an adjustable parameter built into these algorithms, “a sort of dial,” says Knott, which could be turned up or down to build more randomness into the content that the user is delivered. Regulation could require this dial to be turned up, for example, and thereby ensure people are exposed to a wider range of perspectives. The challenge, says Knott, is that this would make the algorithm less good at doing the thing that most users have come to like about it - its ability to deliver content that matches my preferences.

Picking up a similar point, Andrew Butler argues for regulation, but says that the point has to be reached at which we are no longer willing to endure the pain of all the harmful impacts of unregulated digital media in exchange for the convenience it gives us. It comes down to a question of what we are willing to endure for convenience, he says, and “when you’re talking about a high level of convenience, as you are with digital media there has to be some level of pain before people will say you know what actually this convenience is just not worth having.” In relation to digital media, he says, the level of pain that we are not willing to endure for convenience may not be any of the risks it poses to democracy. It may instead be the harmful effect we see digital media is having on our children. “Is the pain in the area of digital, that we’re getting kids who are getting addicted to digital stuff?”

Butler suggests that to get people across the line of being willing to trade off a little convenience, we may need to remind them of the role regulation has played in establishing the social infrastructure we have at the moment. “[I]t’s not just something that you can take for granted, but of course human beings can’t help but take the status quo as being for granted.

Even if we agree that regulation ought to be pursued, we face a second question. Will it work? Paul Moon says no. “Regulation doesn’t work,” he says, “It has the opposite effect. Every time something is banned, or prohibited, or discouraged, it creates an attraction to it.” As an example he says the efforts to prevent Lauren Southern and Stephen Molyneux only drew more attention to their planned visit and their message. “You get a ban. Everyone asks ‘What’s this?’ and instead of an audience of a couple of hundred of die-hard supporters, you get hundreds of thousands of people looking at their videos.”

Jurisdictional challenges to regulation often came up in interviewees. People identified difficulties enforcing domestic standards of, for example, harmful communications, when the sites hosting those communications could be hosted anywhere in the world. A common concern was that the tech giants, in particular, would simply ignore regulation. However even smaller operators, who did a small portion of their business in New Zealand, could be hard to hold to account if they had no physical presence here at all.

John Hancock, of the New Zealand Human Rights Commission, for example, highlighted the jurisdictional challenges of personal data about New Zealanders being held offshore. “A lot of our information is kept offshore. So for us, that’s an issue. A lot of personal information [belonging to New Zealanders] is in the cloud, a lot of our working information is in the cloud. It’s held in a data server that’s not within New Zealand territorial jurisdiction.”

Erika Pearson argues that these multinational jurisdictional challenges are a reason for regulators and legislators to ensure that they understand exactly how digital media, and the infrastructure that support them, actually work. At the heart of this challenge, says Pearson, is the question: “Why is online abuse legally different to threats which people would be arrested for in real life?” The answer generally given, she says, is jurisdictional. “Because someone can be sitting in Miramar, but because [their communication] goes to a server in Los Angeles and get bounced off a thing and it lands in a computer somewhere else— that transition somehow de-contextualises it from our current legal system.” This is not a good argument against regulation, says Pearson, but it is why “I insist that it is really important that we should understand systems.” These issues are examined in more detail in the section on jurisdiction below.

The speed and scale at which digital media could distribute content was also identified as a challenge to regulators, with the video made by the perpetrator of the Christchurch mosque massacre often given as an example. Because digital media can be easily replicated and shared, it can be very difficult to locate all copies of content that breaches legal or regulatory standards. Pornography was another example commonly given to illustrate the challenge of regulating online content. There are ways, people said, to get around most kinds of regulatory barriers - including age or identity verification.

But the fact that regulation works imperfectly, says Andrew Butler, is not a good argument against it. Regulation, he says, plays an important role in establishing social norms, and that role can be effective without perfect implementation. A good comparison, says Butler, is speed limits.

“We have a law that says the speed limit is 100 kilometres an hour. Does the law mean people don’t exceed the speed limit? No. Does the law mean that most people who speed get caught? No. But what it does, it fixes a community standard.” These community standards, Butler says, are upheld not only through legal enforcement but through social enforcement. “If you speed, your kids are in the back of the car saying ‘Dad, your breaching the speed limit, I don’t feel safe.’” So asking whether people obey a law 100% of the time is not a helpful way to measure the effectiveness of law or regulation. “That’s not how law works,” says Butler, or at least, “that’s not the only way in which law works.”

Child rights advocate Sarah Morris agrees that neither total obedience nor full accountability are the right measures of effectiveness of regulation. Regulation can have a protective effect on child rights without being perfectly implemented, she says. She gives the example of the time bands for broadcast programming and advertising, which are designed to ensure that people know that the content being broadcast before a certain hour is suitable for children. Despite the availability of YouTube, and on demand television services, Morris says “those time bands are still incredibly important to protect children from viewing inappropriate content. The idea that children can see anything they want to now, on demand, doesn’t mean that we shouldn’t still have regulations and controls to protect them.” It means that we have to find new ways to establish clear social standards,

including via regulatory controls, around which content is suitable for children on digital media.

Inadequacy of existing regulatory environment

Across a range of different areas, interviewees noted that the current regulatory environment is not up to scratch.

In relation to the protection of privacy, one of the challenges of the current regulatory environment, is its focus on harm, says John Edwards. “One of the big difficulties I had in applying our regulatory framework to the Cambridge Analytica situation is that we have a harm-based enforcement regime.” What that means is that our Privacy Commissioner can’t proactively go to a company using information for a purpose other than the purpose it was collected for, and require them to stop doing so. “I have to wait until someone comes to me,” says Edwards. That person needs to make a complaint, which requires them to understand something about how the Privacy Act works, and they also need to be able to establish a harm to them arising from that breach. They need to be able to show, says Edwards, that they have suffered actual loss or significant injury to their feelings, or a significant indignity.

A common criticism of our current regulatory environment is that it relies on proactive complaints by users. This in turn relies on users having some idea not only of the rights and protections afforded to them by our regulatory framework, but also which official entity they should go to for redress if those rights have been breached. All of that depends on users having the time, ability and confidence to pursue a complaint, which means that people who are already at any kind of social or economic disadvantage - in other words those people who are more likely to be the victims of online abuse - are further disadvantaged in their ability to pursue any kind of protection or redress.

Another commonly identified weakness our current regulatory environment is in the area of artificial intelligence. Some interviewees raised concerns about the use of AI in government, for example in predictive policing, and the inability of the current environment to regulate this. While the use of AI in government falls outside of the focus of this research, some of the issues it raises overlap with the issues raised by digital media, and as Alistair Knott points out, working out how to regulate the use of AI in government decision-making may be a good starting point for working out how to regulate the use of AI by digital media platforms. In relation to his own research, Knott says, “we were going to look at the use of [AI] in government first because ... it is just like an easy target. Whereas, asking how to regulate Google or Facebook’s use of these things is much harder.”

Jurisdictional and legal issues

Amongst the majority of interviewees who argued in favour of a regulatory response from the New Zealand government to the risks and threats posed to our democracy by digital media, there was a common recognition that this regulation would need to happen both at a domestic or national level, and at an international or multi-state level.

Regulation at the multi-state level

Economist Ganesh Nana emphasises the need to tackle the challenges posed by digital media at an international level. He gives the example of multilateral efforts to find ways to effectively and fairly tax multinational corporations. “The OECD and IMF are grappling with this. How do we tax these institutions across borders? Because society come to the agreement that it is just not fair how these big corporations can choose to reside their activities in low tax environments and get away with it.” The way forward, he suggests is likely to be found in a combination of citizens and users demanding a fairer regime, and some form of international cooperation. Although he also sees “quite clearly that New Zealand on its own, passing some regulations may be helpful in some instances.”

Abiding by our regulatory standards should be a requirement to do business in our country, said one participant. “[I]f your product is going to circulate in our country, and be available to our people, here’s how we do it in our country.” And New Zealand isn’t an outlier in terms of the sorts of standards we would want to apply to the operations of the tech companies that operate here. “We’re not on our own, we are not the only liberal western democracy that believes in [the risks posed by unregulated digital media] being bad things. There are many countries that are like us. So why aren’t we clubbing together?” If similarly minded liberal western democracies clubbed together, said one participant, a lot could be achieved.

John Edwards agrees that there needs to be some attempt to bring uniformity to the international legal approach to the digital media companies. “Otherwise there’s this regulatory arbitrage. We had this issue with Facebook. We said, ‘You’re here, you’re hoovering up the personal data of two and a half million New Zealanders, you’re collecting advertising revenue from New Zealand businesses. There is very much a presence in New Zealand that is touching on many aspects of the New Zealand economy, but purportedly without having to be troubled by the legal framework that all the other competing businesses are acting in.’ So we need to establish some alignment of legal frameworks across jurisdictions, he says, and “having similar penalty regimes is a really important part of that.”

This is especially important to address the issue of companies “jumping from one jurisdiction to another,” says Erika Pearson. “They have invested all this money in creating huge data farms in, say, Oregon ... but the second that becomes regulatorily unfriendly, you start to see them moving things to whichever framework works best for that dataset.” You can already see this, says Pearson, with Facebook’s efforts to avoid taxation. “Their cloud is legally registered in a third place, like Ireland, to get around tax laws. So, they pay dollars of tax rather than the millions that they should.”

Ian Apperley says companies are constantly looking for new ways to avoid regulatory controls and taxation including by ‘seasteading’ which refers to the idea of moving a company’s data and legal infrastructure into international waters, in order to avoid being bound by any national laws. “[They’ll] just create floating islands. [They’ll] put data centres on there.”

Thomas Beagle sees a risk in multilateral regulation. “It sounds wonderful but it’s also one of the things that really scares me, because I believe that there’s a good chance that ... in twenty years time we’ll have a United Nations Declaration on how we behave on the internet, and I see that as being strongly influenced by both China and the States. And when you say that, you think ‘that may not be that great, actually’. We may be unlucky enough to get what we asked for.”

InternetNZ's Jordan Carter's concerns about multilateral processes for internet governance have already been set out above, and also apply here, in the context of regulation. His argument would be for a multi-stakeholder, rather than a multi-state, approach to regulation i.e. with the participation of the tech companies themselves. Erika Pearson, amongst others, argues that strong representation of citizens, of internet users, would be critical to democracy, fairness and the long term success of any such process.

Regulation at the state level

Alongside this international multi-stakeholder process, interviewees argued for more regulatory and legislative action at a domestic level in New Zealand. As economist Shamubeel Eaqub puts it, "These companies are not special in any way. They're multinationals operating in lots of different areas, doing lots of different things." So while international standards ought to be pursued to set what Eaqub calls "meta-rules for the big things", like disinformation and abusive behaviour. Those global standards would set the minimum, but New Zealand can "choose to have a higher standard here."

Joy Liddicoat argues that we might need to consider even stronger protections to guarantee the rights of New Zealanders in digital spaces, perhaps even through constitutional protection.

"We don't have any constitutional protection of access to the internet," says Liddicoat. "We don't have an overarching right to privacy ... We actually have very little regulation. Encryption is not outlawed. Anyone can encrypt anything here. It does need more exploring."

Regulation on specific areas

Certain specific areas tended to be suggested when interviewees made the case for better regulatory protection in New Zealand. These included antitrust and competition, privacy and use of data, disinformation and deep fakes, political advertising and micro-targeting, and online abuse including hate speech.

Antitrust and competition laws

At the most structural level, interviewees pointed to the need for New Zealand's antitrust and competition laws and regulatory systems to be reviewed and updated to keep up with the tech giants. "There are these global monopolies," says John Edwards, "and I don't think antitrust law is yet suitable, or has yet figured out a way to deal with it."

In order to adequately regulate these companies, one participant says, we need to work out which standards we expect them to comply. "I think there is a good argument to say these businesses have not done enough, but I would then quite quickly say, measured against what? What are we expecting them to actually achieve?" These companies are not the same as traditional media companies. So we need some agreed definition of where they fit in our regulatory systems, and what standards are therefore applicable.

The question of which standards appropriately apply to digital media companies was also addressed by Dr Every-Palmer. He argues that the conventional antitrust framework can adapt to digital media markets. He gives the example of the proposed merger between

Stuff and NZME, which owns the Herald. “In this case, the Commerce Commission said, ‘OK, there’s the issue of media diversity here, because Stuff and the Herald are so big and would reduce media plurality if they combined.’ In the end, he observes, the Commission reached the view that “despite there being potentially up to 200 million dollars of benefit in terms of efficiencies (cost savings), this was outweighed by the negative impacts on media quality, media diversity and democracy. And so it said no to the merger.” This was upheld by the High Court and the Court of Appeal (Dr Every-Palmer acted for the Commission).

Whether or not you agree with the conclusion reached in that case, Every-Palmer says, you can see that the Commission had a clear picture of the role of these media companies in our democracy, and the democratic importance of diversity, and the antitrust framework could adapt.

However, he notes that social media platforms give rise to many regulatory issues that sit outside the conventional antitrust framework. “[I]t’s the underlying issue throughout,” says Every-Palmer, “what do those businesses do? They do not produce original journalism, but they are an important distribution channel. Facebook users come across news stories that have been selected by an algorithm. They may also come across fake news that has been paid for, as we have seen with the US Presidential election and Brexit. Given the importance of the news media for the democratic process there are very real issues here that we are only just starting to grapple with.” Although the antitrust framework can promote competition which may ease some of these problems, he does not consider that the antitrust framework can be a sufficient solution by itself.

Shamubeel Equb agrees that there are challenges in regulating digital platform companies, but points out that this is not a unique challenge. “Quite often we’ve seen this, where there are these difficult to regulate industries, but it is possible to have global coordination, and it’s possible to have minimum standards, and by and large have worked pretty well.”

Equb gives the example of banking regulation which used to be, he says, “really fragmented, and national.” Then, following the Second World War and in response to some of the risks of that fragmentation that were shown up during and in the aftermath of the war and Great Depression, there was a lot of international coordination in an effort to settle on some form of international banking standards. “By the 1980’s we had the Basel Rules, from the Bank of International Settlements. Which is a kind of broad consensus of minimum standards that everybody should follow.” Not everybody follows them, he acknowledges, but they set a standard that you could “hang your hat on, it was this line in the sand, that was really useful.”

Privacy and data use

Privacy is another area which was often highlighted by interviewees as needing better regulation in the digital age. James Every Palmer illustrated the urgency of this issue by posing a hypothetical scenario. “If the government tried to pass a law requiring us to carry a device attached to us at all times, and recorded information through that device and sent it overseas to a foreign company for uses we weren’t aware of, people would probably be marching in the streets.” Instead, he says, most people are voluntarily paying large amounts of money to do exactly that, but for a privately owned company which isn’t accountable to us in anyway. “[I]t’ll be interesting how history judges this period of time in terms of our naivety,” says Every Palmer, “in terms of the way we’ve thought of digital technologies as being a neutral platform, just a market development, not worthy of the same amount of regulation that we would give to an electricity distribution business.”

Specifically in relation to privacy protections, the Office of the Privacy Commissioner made several recommendations on the review of the Privacy Act in 2018. The interview for this research took place after the Privacy Commissioner had made his submissions on the review, but before the Select Committee had reported back. At that time, John Edwards explained the areas that his office thought New Zealand's privacy regulation needed to include, which were not covered in the existing legislation. They included data portability, and algorithmic transparency, which included "the right to have transparency over automated decision making and the right to object to automated decision making." They also suggested aligning New Zealand's privacy law with other data protection regimes around the world.

On March 2019, the Justice Select Committee reported back on their review of the Privacy Bill. The Office of the Privacy Commissioner welcomed their report saying, "the Committee has listened to submitters and the reported back Bill contains measures to ensure the law addresses some of the most pressing aspects of the modern digital economy." Of particular relevance to this research, the law as will now explicitly set out that when agencies collect, process and use New Zealanders' personal information they have to comply with New Zealand law, "regardless of whether these agencies have a brick-and-mortar presence on our islands." The Bill also limits the exemption given to news media to only those media that are subject to the oversight of an appropriate regulatory body. The Bill doesn't include everything the Privacy Commissioner had sought, and in a public statement John Edwards said he will "continue to make the case for more civil enforcement powers and other modernising reforms to ensure that New Zealand's privacy framework is robust, fit-for-purpose and comparable to those of its trading partners."

Privacy expert, Marcin Betkier agrees with Edwards that our current regulatory environment is inadequate to respond to the changes brought about by digital media. One option, he says, would be for New Zealand to incorporate the same framework as the European Union's GDPR although he raises concerns, as outlined above, about the weaknesses of a consent-based model. In the end, Betkier argues, we'll need more than an updated privacy law to manage the challenges of data use and consent in the digital age. His proposals for new structural solutions to those challenges are set out below.

Disinformation and deep fakes

In terms of disinformation, Andrew Geddis argues, we need to find a way to create some sort of disincentive for the platforms. "We've got to find a way whereby spreading disinformation can have a potential cost to it that would cause them to at least stop and ask – 'Whoa, is this going to be worth it?'" One cost you can attach is the risk of legal liability, and Geddis offers the parallel of environment regulation which creates a disincentive for pollution. "So, if you are pumping shit out into the stream, [there is a risk that] you are going to get fined."

Another way to create a disincentive, says Geddis, is to attach a reputational cost. The parallel for this is ethical consuming, where people know that a product or service is made unethically which puts them off purchasing or using that product. Geddis suspects that reputational risk will be most effective in creating an incentive for the platforms to monitor and prevent disinformation campaigns. "But [that] gets back to the need to make sure that there is a tangible name attached to such campaigns in a way that can then be tracked." This would require some sort of identity verification by the platforms, which is

discussed in more detail below.

Tom Barraclough and Curtis Barnes have formulated and adopted a framework approach to synthetic media and deepfakes that can be applied in an iterative way to cope with the pace of new development in this area. Using their framework, they identify an existing body of New Zealand law which already deals with the potential harms of synthetic media. This body of law recognises an important balancing of factors between individual privacy, freedom of expression and the rights of creators. They note that the boundaries as drawn in existing law should be carefully considered before moving to implement new law.

But they also identify pragmatic challenges to enforcing that existing law. These include: difficulty enforcing it through and against international platforms; barriers to access to justice; the speed of harms online compared to the pace of legal response; and a global limitation on access to evidential services to assess the reliability of audiovisual information.

Political advertising and microtargeting

Identity verification also comes up when interviewees talk about options for regulating political advertising. James Maclaurin goes as far as to suggest that political parties shouldn't be allowed to micro-target. "I think it would be better to make micro-targeting illegal for political parties in New Zealand. Now, there might be ways of getting around that, but there would be a real risk for parties that tried to do so. I don't see that there would be very much lost if you did this."

Andrew Burns disagrees, saying that targeted advertising on digital media has made political campaigning more accessible, and specifically more affordable, for independents and candidates from smaller parties with less money. He argues that 'banning political parties from targeting people on social media, could actually result in a degraded democratic environment.' He believes 'you couldn't really ban 'micro-targeting' without actually removing many, if not most, targeting options, and that a blanket ban on targeted advertising is the wrong way to go. He instead advocates for better verification of the identity of political advertisers and more transparency so that people being targeted can easily see who paid for the advertisement they are seeing and why they have been targeted.

Burns argues that banning political parties from targeting people on social media (because you couldn't really ban 'micro-targeting' without actually removing many—if not most—targeting options) could actually result in a degraded democratic environment. "The reality is that a majority of people get their information and news online and from social media," he says, "If you effectively ban political parties from presenting policy and viewpoints in the place where the majority of people get their news, then the most important actors in the political process, political parties, are unable to contribute to political debate." It would be like banning political parties from printing newspaper ads or placing broadcast ads during the election period. Social media companies have gone a long way to restrict the misuse of their platforms, says Burns. "In fact, they've gone so far it makes legitimate use sometimes difficult – that's not to say there isn't more to be done, but a blunt instrument that effectively excluded political actors from these advertising platforms is the wrong way to go."

Nat Dudley points out that Facebook already does require some validation of identity

CHANGE CONTEXT & ENVIRONMENT

for advertisers, requiring proof of address and information about the legal status of the organisation in the form of, for example, a company's registration. "So they do already play in the field of authenticating you are who you say you are. It's probably not a massive stretch for them to do more of that."

Dudley's view is that Facebook's willingness to require verification would probably not stretch to denying the right to advertise to unverified accounts. "I think if you asked them not to sell advertisements to someone they hadn't validated they'd say 'Throwing away money? Lol no.'"

A very strong incentive, possibly in the form of significant fines, would be required to dissuade a company like Facebook from selling an advertisement to someone who was willing to pay for it. It might be easier, says Dudley, to get them to make it very obvious to the person see the advertisement that it was from an unauthenticated source."

Another regulatory solution raised in the context of micro-targeted political advertisements was the potential for a centralised 'audit' function to be performed on micro-targeting algorithms to ensure they are accurate and sufficiently transparent. Andrew Chen, PhD candidate at the University of Auckland working on computer-vision and privacy, argues that "there should be some trusted body, a government body or similar, that can come in and check that these algorithms are actually doing what they're supposed to do, and that they are following all of the relevant regulations."

He points out that there is an additional need for a central auditor because of the 'black box' nature of these algorithms; "you chuck a bunch of inputs, you've designed the system to learn, you get your outputs out at the other end, no one really knows what happened in between." An audit function, he argues, would allow for some expert regulatory oversight over the content of micro-targeting algorithms without compromising trade secrets within commercial companies.

The next broad category of responses proposed by interviewees are those which attempt to change the context and environment within which digital media operates. These include proposals for new initiatives which would either supplant digital platforms with publicly funded alternatives or support the growth of competitive alternatives. This section also includes proposals for responses by the platforms themselves, collective actions which could be taken by the employees of those platforms and by their users, and mechanisms for greater monitoring, accountability and transparency by the existing platforms and any new platforms as they emerge.

NEW PUBLIC INITIATIVES AND COMPETITIVE ALTERNATIVES

Most of the suggestions made for new public initiatives related either to more public support, including funding, for public interest media or to the creation of a publicly owned and democratically accountable infrastructure for digital democracy. A third category of suggestion was for new public principles, frameworks and structures that would better protect the rights of citizens in digital spaces.

Public interest media

Many interviewees agreed that public interest media and journalism have a big role to play in a functioning democracy - both in terms of ensuring that citizens are well-informed on matters of public interest and in terms of creating some shared understanding - a 'democratic consensus' - on what those key issues are. Having recognised this, and the challenges posed to traditional funding models for public interest media by the growth of digital media, some interviewees suggested more public funding for public interest journalism and media was needed.

"Perhaps the answer is to make more public funding available," says James Every Palmer. "I guess one model is to beef up Radio NZ so it's more like BBC. But also making funding available on a contestable basis avoids some of the potential pitfalls from having a state broadcaster having such a big role as the BBC does." While he says the BBC generally does it well, Every Palmer cautioned that "in most regimes, state funded broadcasting is more problematic than beneficial." Recognising that it could be harder to measure key performance indicators is there is more contestable public funds for public interest media being made available to a wider range of media producers was. He nevertheless was of the view that it would be one way to improve public interest media without the risk of too much state control.

New models of funding for privately owned media platforms were also seen by some interviewees to offer promise as a means for providing financial viability in the future. These include the use of paywalls and subscriptions, premium subscriptions with added-value services like a daily news analysis, and voluntary digital micropayments for specific articles. All of these options have challenges, but as digital payment options become more and more efficient some of those challenges may be reduced. As Every Palmer says: "[I]f micropayments develop, so it becomes more seamless for a website to charge a cent or a fraction of a cent per page viewed, you can imagine that that's something that could work for a news business. So it could be that they become more profitable again."

Publicly owned spaces for digital democracy

Another area where interviewees saw a role for public funding was in the establishment of a publicly owned, transparent and democratically accountable digital infrastructure for democratic purposes.

The big digital platforms were not built with the goal of providing an infrastructure for democracy. They have become spaces in which core features of our democracy are enacted, but without that purpose being built into their design, business model or governance. One approach to this is to try to make the platform companies change their designs and governance to be more transparent, more diverse and more accountable. But another approach is to accept that the core functions of democracy cannot be provided by virtual monopoly private corporations, and to plan instead to create a public infrastructure for that purpose.

Pia Andrews argues that government has a responsibility to build a digital public infrastructure. "Government has been trying to outsource, and outsource, and outsource its responsibilities. At what point, and I think that point is now, before it gets too much further, does government needs to step up and say 'What is the digital public infrastructure that's required for an effective, fair, socially inclusive society?' The

tech giants, she says, have built these platforms in a vacuum of government response to people's needs. This is not an argument, she says, that government 'should run all the things.' But, she says "government should understand and support digital public infrastructure. Where it is genuinely left to the market, governments should at the least put some upwards pressure, so a minimum standard of quality and of rights is maintained."

Investment in research and development

A common suggestion was for more investment by government in research and development in these areas, including research to test some of the solutions proposed in these interviews and investment in developing alternatives to the privately owned digital media platforms.

Some participants also suggested areas for further research within their specific field of expertise. Tom Barraclough and Curtis Barnes, for example, identified future research on the impacts of synthetic media including:

- Refining their framework and applying it to other legal systems preparing for the impacts of synthetic media.
- Creating law and policy systems that allow for a balance to be struck between the rights of creators, consumers and citizens to privacy, freedom of expression, and intellectual property.
- Methods of effectively preparing and insulating individuals and institutions to critically analyse audiovisual information before relying on it in ways that create a risk of harm.

New principles, frameworks and structures

Along with proposals to build a public infrastructure for digital democracy, there were proposals to establish new legislative frameworks to better protect digital rights, and along with them, new structures to monitor and oversee the implementation of those frameworks. Parallels were drawn to the existing framework for human rights, which has the Human Rights Commission - a publicly funded but independent crown entity - as a monitor.

One of the specific ideas proposed is a Digital Bill of Rights, with an appropriate watchdog to monitor the implementation of those rights. Pia Andrews argues that the ability of the big tech companies to get around legislative protections by creating terms and conditions that impede on the rights of their users is why "digital rights is so critical". It's also why "transparency and accountability is so critical, which is why governments role as a watchdog is critical."

James Every Palmer likes the idea of a digital bill of rights, but she says the process used to develop it would be critical and proposes some sort of Technology Commission to lead the process. The Commission would be made up of people with "core expertise" whose role would be to ask what the core principles should be for a digital media platform. "What do we expect it to do? To control content, to control bullying, to control fake news, to allow people to switch to competing social media networks?"

This kind of Tech Commission wouldn't, he says, come up with easy answers to any of those questions but "at least there'll be someone trying to carry along the dialogue about

them, and more likely to lead up to a joined up conversation about it, and it would also be somewhere where people with a range of concerns about the modern online world would be able to bring them.”

Some sort of monitoring body along these lines could be a way to mitigate against the previously identified challenge of people needing to somehow work out which of the myriad entities with some responsibility over digital spaces is relevant to their specific concern. If it were designed and resourced to have a proactive monitoring, research and advocacy function, this kind of body could also help address the power imbalance inherent in the current system, which requires people whose rights are being violated to proactively identify this and raise a complaint.

One concern about this sort of centralised response, says Every Palmer, is that it would put the technical expertise needed to respond to the challenges of digital media into a silo. “[A]ll of government are impacted by online technology, so if we create that centralised expertise, won’t that mean that we take the experts away from the other places where they’re needed as well? And would that give those other places an excuse not to do anything about it?” Every Palmer thinks these are valid concerns, and certainly they echo concerns that have been raised about the impact, for example, of centralising human rights expertise given the need for human rights concerns to be engaged and addressed across all areas of government policy and practice. But even considering those concerns, he comes out in favour of having a centralised body of expertise.”

These proposals get to the heart of one of the key challenges identified by Jordan Carter and other interviewees - the need for more ‘joined-up’ government thinking about the internet generally, and about digital media specifically. This challenge can be seen, at a superficial level, in the way that Ministerial portfolios are allocated - so that digital media is included in the broadcasting portfolio, digital government is included in the open government portfolio, and online abuse and harm sits within the justice portfolio. In some ways this is good, it makes it clear that digital technology can’t be siloed. It has implications across many, perhaps all, other portfolios. On the other hand, it can create challenges for thinking about how all these different issues - ranging from online abuse and misinformation to cybersecurity and election advertising - are connected through the underlying structural issues posed by, for example, the governance structures, business models and anti-competitive practices of the tech giants.

Carter gives the example of encryption policy, which he says “isn’t a national security and law enforcement issue, [it] is a whole of economy and society issue.” In 2017, InternetNZ produced a paper outlining some of the policy issues raised by encryption. The paper called for ‘a whole of economy discussion,’ on the issue. Carter requested meetings with a range of ministers with responsibilities for sectors that were impacted by encryption, whose offices responded by referring the matter to the Minister of Communications. We faced a similar challenge trying to secure interviews for this research project. Requests to speak to policy experts across the full range of issues raised by the research were repeatedly referred to the office of the Minister for Communications. While digital media clearly raises substantial issues in the area of broadcasting and communications, our attempts to point out that it also raises policy issues across a number of other portfolios seemed to fail. As Carter says, “they don’t even know that they don’t have joined up thinking. So, that’s a risk.”

Some interviewees pointed out that efforts to appoint a Chief Technology Officer may have been an attempt to address this lack of joined-up thinking on digital technology across government. The difficulties of finding in one person the full range of expertise and experience needed to help fill those gaps perhaps illustrates the need for a team approach. For one thing, if the government is to avoid repeating the mistakes of Silicon Valley, they'll want to be sure that the people put in charge of coordinating government responses to digital technology represent a more diverse spectrum of the population.

The establishment of a Digital Economy and Digital Inclusion Ministerial Advisory Group, several members of which were interviewed for this research, may be a step in the direction of a more joined-up, diverse leadership for government responses to the opportunities and challenges posed by digital technology. On the other hand, the Prime Minister's consultation with an unnamed group of 'advisors' in relation to her approach to the digital platforms in the wake of the Christchurch mosque massacres could be seen as a step away from the sort of transparency that is so critically needed in this space.

Case study for a new institution: Privacy Management

Marcin Betkier proposed a very specific new kind of institution to better manage and protect people's privacy. Technologies that read our thoughts are in development, Betkier says, and technologies that track our movements, our conversations and our activities at home are all already widespread. We are already giving away more information - more knowledge - about our intimate lives than we probably realise. So, a new approach is needed, he says, as technology gets "closer to us, and closer to our bodies and minds. [W]e need new barriers. We have to create the equivalent of [a door that we can close] if we want to have the similar organisation as a society as we are, as we have right now." This new kind of door, says Betkier, will need to be created through digital technology, because the "world is moving towards more widespread, more intensive use of technology." We can't realistically stop using technology, he says, so we'll have to create a kind of barrier that is under our autonomous control. Something which protects us from being exposed to technology because otherwise "we'll be fully readable".

Betskier proposed a new institution - independent of the corporations and public sector organisations who want to use our data. This 'middle space' would enable people to manage their online privacy. He calls it a middle space because it falls between two extremes on a spectrum of data centralisation. At one end you have the current situation where all data sits with the service providers, whether they are private corporations or public services; this data is fully visible only to them and only they know how it is really used. At the other end of the spectrum, all of that data would be held and controlled by individuals, which means that services that might be able to use that data for public good, or to provide services that people want to benefit from, wouldn't be able to fully develop. Individuals also lack the expertise in using and securing their own data. So, Betskier's proposal is "in the middle" of those options - to initially leave data where it is, but to regulate technical and organisational mechanisms that bring all this data under the full control of individuals.

This new middle space would be occupied by third parties - he calls them Privacy Information Administrators - that work for individuals and do not benefit from the use of personal data. They would be acting as a watchdog and facilitator, helping people manage their data according to their personal values and preferences. One way that could work

would be by helping people proactively think through the choices and trade-offs involved in sharing their data with a company or government department. This could be done through, for example, an online survey. Once the individual has established their personal ‘privacy profile’ in this way, then the Privacy Information Administrator would act as the authoritative source of privacy settings for all entities wanting to collect and use people’s data. If my personal privacy profile provided, for example, that I was happy for my health service provider to have access to certain kinds of medical data about me for research purposes, that data could be provided to them without coming back to seek my explicit consent. If, however, an application I were using, like a running app, asked for access to a new kind of data, which wasn’t currently allowed in my privacy profile, they would have to make this request to the Privacy Information Administrator, which would then send me a request explaining how the new data request fell outside of my existing preferences and what are the possible consequences.

This kind of institutional approach could reduce the number of times in a day or week that each individual was asked for explicit consent to share data, and help overcome the problem of consent fatigue identified by several interviewees in this research. Furthermore, Betkier proposes technical and organisational mechanisms that would put the collection and use of personal data by the data controller’s under the monitoring of the Privacy Information Administrators. Overall, in this proposal individuals would have a knowledge what data about them is used and how it is used, and the ability to act by changing their privacy settings and/or deleting their data.²

Erika Pearson also sees some potential in this kind of institutional model, to restore control of personal data to people. “People who have opted into Facebook at that point of consent, they have no recourse to manage their data. I have seen a few models that [address that problem], and one is that institutional model.” Similarly to Betkier, Pearson draws a parallel between data and bodily sovereignty; “So, you have sovereignty over your organs and they can’t harvest you without your consent. But you can give blood. And so, the idea is that data is a part of me and I am giving this to you for this benefit, but I can pull it back at any time and you have to tell me when you change the use.”

Importance of inclusive and collaborative approaches

Through all these proposals for new institutions, new frameworks, new infrastructure was woven a consistent message about the importance of process. Whatever process is used to develop these kinds of responses, it needs to be transparent, inclusive and democratic. As Joy Liddicoat pointed out, a big part of the problem of having corporations like Facebook and Google control important parts of our democratic infrastructure is that they have no obligation to engage in any kind of democracy. “As a user I don’t have any say, I’m never polled on ‘What should our privacy policy be?’ ... We have no input whatsoever into their guidelines, or standards, we’re always arguing about it from the outside, in multi-stakeholder forums.” Liddicoat sees an opportunity in New Zealand “to say to some of some of the corporates here, ‘Okay, step up,’ and engage users in creating the online standards.”

Similarly, government responses need to be genuinely collaborative. As one interviewee put it, “You’re not going to get the results with a co-design process which involves only government and industry that doesn’t engage with the people worst affected by these platforms.”

2. The full account of Betkier’s proposal and its economic implications is available in his forthcoming book: *Effective Privacy Management for Internet Services: Economic, Technological, and Legal Regulations* (Intersentia, Cambridge, June 2019).

PLATFORM RESPONSES

The major platforms companies have responded in many ways to the challenges identified by interviewees in this research. It's not surprising, given the nature of digital media, that those responses are often iterative, even experimental, and that they change regularly. Given the pace of that change, early attempts to include a summary of the most recent platforms responses into this research were abandoned in favor of addressing those responses when all the other analysis has been completed. To that end, the major platform companies will be presented with the overall findings of this research and given the opportunity to outline their existing and planned responses.

This section is not, therefore, a summary of existing responses by the platforms, but instead outlines suggestions made by interviewees for ways in which they could respond to the opportunities and challenges identified. Most of those suggestions fall into the category of platform design and affordances, and cover a range of design issues including algorithmic transparency, identity verification and content moderation.

Governance and management

Before we get into the discussion of platform design, a few interviewees identified changes that could be made at a more structural level by the platforms. Mostly these related to the diversity and inclusiveness of the governance and management of the companies. More diverse leadership, some argued, would better equip these companies to meet the needs - and protect the safety - of their extremely diverse users. Erika Pearson suggested that a more inclusive, and less 'paternalistic' approach from the platforms could lead to better identification of risks and harms to, for example, already marginalised groups. "It would be lovely to see cultural groups have negotiation power. To be able to say okay, we like your platform and we will let you monetise us, but we want these specific things in return. We want Samoan language moderators, or we want the ability to manage two-tier friending behaviours to accommodate cultural norms."

Platform design & affordances

Suggestions for more ethical platform design were common. As Richard MacManus put it, "this whole trend towards ethical technology, try to put some sort of moral sense back into these tools." Many of the leaders of the ethical tech movement have themselves come out of working for the big platforms. Tristan Harris, for example, who set up the Centre for Humane Technology, spent three years as a Google Design Ethicist. As MacManus sees it, this gives them both expertise and influence. "I think their intentions are really good, and they do have influence."

The motivation for the big companies to look at more ethical design, says MacManus, will probably come from pressure from users, and particularly younger users who are more aware of the kinds of risks and threats outlined in this research than their parents were. "The younger generation are not flocking to Facebook as our generation did, so you have to think that [the platforms] will do something to course correct." So far, he says, that course correction has come in the form of buying up competitor products that have proven to be popular with younger users, like Instagram and WhatsApp. MacManus questions whether Facebook's leadership have "quite grasped that they need to move into a more ethical, more humane approach to this technology." Other interviewees

argue, as outlined above, that regulatory pressure could be brought to bear at the national and multinational level to create the necessary motivation for a shift to more ethical design. While others, like Jordan Carter, say that only a multi-stakeholder approach with the active participation of governments, the platform companies and civil society will achieve this.

Algorithmic transparency

One area where the platforms could - and in some cases already have - design solutions to mitigate against the risks identified in this report is through greater algorithmic transparency. Nat Dudley refers to demands that already have been made, for example, for greater transparency about the placement of political advertising. “For you to be able to see, firstly what all of the ads are that have been placed, not just the ones you see, and secondly what targeting criteria they used.”

One of the challenges to algorithmic transparency, says Dudley is that “even within somewhere like Facebook, most people don’t understand how that algorithm works. It’s a complex beast.” But movement is happening, she says, towards allowing people to see what happens to the data that is collected from them, and thereby allowing them to make more informed decisions about whether or not they want to share that data or participate in that exchange. Another area of movement, says Dudley, is towards better transparency about how your data is actually used and how that affects the information and interactions that you see. Both of these movements, however, are ‘still quite individual’ says Dudley, which makes it hard for people to see the systemic implications. “[I]t’s still very hard for you to then imagine, the ongoing effect. Like being able to construct a model that says this is how opinions are shaped and formed. It’s very difficult right now.” But movement towards more transparency both about the data being collected and the way it is used about the building blocks for people to be able to understand what’s going on.

Identity verification

Another commonly suggested area for better design solutions is identity verification. As set out above, there are a range of views as to the risks and benefits of allowing some degree of anonymity or pseudonymity online. Nat Dudley made the case for pseudonymity, rather than requiring real-name verification on the basis that many of the people who depend on pseudonymity online are people for whom participation in public discussions on matters relevant to their lives could otherwise be fraught with risk.

“There have always been places that have played around or investigated essentially cryptographic validation of who you are online without a real name identity associated with them,” says Dudley. Examples include journalists using the Pretty Good Privacy (PGP) encryption program to communicate securely with sources, and tools like Keybase.io, which allows users to “prove” a link between certain online identities (such as a Twitter or Reddit account) and their encryption keys.

All of these tools provide ways to validate that someone is who they say they are, without requiring them to reveal their real name. The biggest problem with those tools, says Dudley, is that the user experience is terrible. So most people either can’t be bothered using them, or when they do, they use them incorrectly. We are starting to see some more usable tools, says Dudley, “like Signal, or even WhatsApp which has encryption built in.” But progress towards widely usable tools has been slow.

James Maclaurin proposes ‘another pole that we could push to’ on verification which he says “would solve some of these problems but make it a very different world.” His proposal is that you could only follow people on Twitter, for example, who are either your actual friends or verified publishers. So “you can follow people on Twitter if you know them. And then there are some people you could follow who you don’t know, but they are the people that have the provenance.” He admits that this is not a solution he’s thought about a lot in terms of how it would be managed in terms of verification, but offers it as an example of a middle ground between requiring all accounts to be verified and leaving things as they are.

Content moderation

Content moderation is another area which attracted a lot of proposals for design solutions. In the wake of the Christchurch mosque attacks, content moderation attracted a lot of attention from the public, in the media and from politicians. One of the risks of this kind of heightened focus on one aspect of this much wider set of challenges is that solutions are proposed which fail to address the many related and connected issues outlined in this report. However, interviewees did make some proposals which were specifically focused on the issue of moderation.

Thomas Beagle says moderation has been key in his experience of forums and online messaging systems. “Moderation has always been the key. I don’t mean moderation of people’s behaviour, I mean someone moderating that forum. You can set the tone quite early.” Beagle gives the example of Reddit forums, which often have the reputation of ‘being a sewer, you know, why would anyone go there?’ But at the same time, he says, he’s part of various Reddit forums including one on DIY and several political ones. “And to be honest they’re all lovely. Any unpleasant behaviour gets shot down and removed very quickly.”

This works, he says, because the forums are on a small scale and within shared communities in which somebody or bodies have taken on the responsibility of maintaining the tone. Twitter, he says, doesn’t work that way. “We don’t come there to communicate. The same with Facebook as well. And moderation is, I’m going to say largely impossible on that scale.” Beagle says he nonetheless does see people asking for better moderation of these big platforms, and giving the example of Germany where strict regulation against Holocaust denial, for example, has required Facebook to employ hundreds of moderators. So he recognises that “we’re smart enough and versatile enough around these things,” to find a solution.

Dave Moskovitz floated the idea, which he pointed out he hadn’t had time to think through in great detail, of using bots to monitor and moderate online conversations. “I think eventually we’ll have bots enforcing civility in our conversations online, and I’m not sure that’s a bad thing. So if I start swearing at someone, or berating them, or using bigoted language to them, the bot will step in and just stop it from going through in real time.” An example of this kind of technology being used in a smaller online community is a bot created to point out sexist language in a Slack channels. If someone sent a message to the entire channel - which included people of diverse genders - and addressed them as ‘guys’, the bot would gently suggest that the user might want to use different language and suggest a range of gender neutral terms.

Moskovitz says there are a range of different ways to approach content moderation in online discussions, but one way or another “we’ll have to develop them because if we don’t, then the internet will fail us, and no-one will ever want to go on.”

Cybersecurity expert Simon Howard says there is a relatively simple solution available to the platforms when harmful content is being uploaded. They can simply turn it off. “Like, when people kept on uploading [the Christchurch mosque massacre] video to YouTube. Why not just turn off uploads for the next 24 hours. You’ve got enough videos on your platform.” He recognises that some people would complain, but given the potential harm, he says, the right thing to do would be to “just turn it off until you get a handle on it.”

Golriz Ghahraman would like to see digital media platforms themselves to take a more proactive approach to monitoring and removing hateful and harmful content. At the moment, the system relies on individual people who are being targeted, to firstly find out that the hateful content exists and then look at it, and engage with it in order to make any kind of complaint. “And suffer the anxiety and depression and all the things that we know comes with online abuse, and then report it.” If you’re running a website, says Ghahraman, “and someone seems to be threatening violence, or perpetrating abuse, surely you have an obligation to report that to the proper authorities yourself?”

One of the notable changes made in the wake of the Christchurch mosque attacks is that some local digital media sites have revisited their practices for managing comments. Stuff, in particular, have identified a list of topics that tend to attract the most problematic comments on their site and will no longer allow any comments on articles about those topics.

COLLECTIVE ACTION

Interviewees recognised potential for both employees of tech companies and users of digital media to use their collective power to demand and create change from and within the big platforms.

Employee collective action

One area in which interviewees observed interesting possibilities was in collective action from tech workers, to either create change themselves or advocate for change from management. There’s nothing new in this idea. At its simplest, it’s an argument for some form of union or collective movement within digital technology. Some of the ethical design movements mentioned above have been started by and organised amongst workers and former workers at the major platform companies.

The need for collective action is highlighted by Marcin Betkier who draws on his own experience working for 16 years in technology companies to explain the challenge. As long as the corporations themselves remain focused on short term financial gain, it can be hard for workers to make ethical choices, in design or otherwise. “[P]eople are ethical but the machinery of the corporation is ruthless.” The focus, he says, is on short term financial goals. “What’s in the next quarter? What’s the result? It’s ruthless. It forces those ethical people to make choices which have little in common with what they would normally choose.”

What makes the possibility of collective action by workers particularly interesting to some interviewees is the combination of the specialist expertise held by many of these employees – their skills lie at the heart of the competitive advantages enjoyed by the platforms – and their relative negotiating power, due to a competitive employment

environment for people with their skills.

On the other hand, as some interviewees explained, many even very senior workers in digital technology are hired on relatively tenuous contracts, relative to the protections afforded by labour law in a country like New Zealand, at least. And whilst their skills may be in demand, there are also always new people coming through with similar skills, and a hunger to work at the big platform companies.

Nat Dudley refers to the work of technology analyst Anil Dash and others like him “who are dissecting the venture capital model and looking what’s going on. What are the motivations? How does that impact the ethical decisions these organisations make?” Anil and others have asked how people – both workers and users – can be empowered to push back against the drivers of the venture capital model. And, says Dudley, they’ve looked at options like a union model for technical employees to give them more power to push back. “Because many of them are ethically uncomfortable with certain things they’re being asked to do,” she says. But the challenge is that people are being paid very well to overcome that discomfort, “and there really aren’t many other places in the world where you can earn those stupid amounts of money for doing comparatively little.”

Collective action by users

Another form of collective action highlighted by interviewees was the potential for change to be driven by organised movements of digital media users. John Edwards says that he can already see a pushback from users, who are demanding more transparency from the platforms. For a long time, he said, the platforms were making changes to data use which people hadn’t fully understood. But over the past year or more, there has been a lot more media coverage of some of the breaches of data privacy, and of practices which surprised and alarmed users, including high profile cases like Cambridge Analytica. This has added up to more user awareness, says Edwards, “and what we are seeing now, is a backlash demand for greater transparency.”

Erika Pearson sees some potential to redress the considerable power imbalance through collective action at a user, or more importantly, citizen level. “There’s this opacity and this power imbalance which comes back to the point about some kind of union or movement to push back. Whether that’s at a cultural level or a national level. We often talk about the nation state in terms of the state side of that, but we probably need to talk about the nation side too.” One example of what that could look like would be a movement for ethical tech which created and implemented an independent system for accrediting digital media companies on their ethics. “An ethical tick for privacy on tech platforms. It would be cool to see the social license account for that and have people actually ask the question ‘What is your ethical process?’ Like we do for clothing.”

LONG LASTING PROTECTIONS

In this section we've gathered the solutions proposed by interviewees with a focus on creating and implementing long lasting protections for citizens and for democracy. One of the recurrent themes in this category are protections against online hate, abuse and harassment.

Whenever the matter of regulating hateful and abusive speech online comes up for public debate, there is a predictable response from certain quarters that this amounts to state censorship. But as one participant pointed out, there have always been limits on speech and publication in New Zealand. "Of course we've had state censorship for years in New Zealand," they say, "we never called it that but that's what we've had. We've had limits on electoral advertising, for example."

Another common argument against the regulation of hate speech and abuse is that the harmful speech will simply be pushed underground. You can't get rid of it, is the argument, so by banning it, you simply push it into some dark recesses of the web. Simon Howard doesn't see what the problem is with that. Regulation should be able to push harmful and hateful content from easily accessible forums like 4chan onto the 'dark web', he says, making it more difficult for ordinary people to access. The dark web, he points out, isn't a place his children are likely to stumble across accidentally. It's not a place that the people who are being targeted by hateful communications are likely to be forced to read them. The darker the corner that kind of speech is forced into, the better, he says.

The Human Rights Commission has advocated for a review of our regulatory and legislative protections against hate speech. This could involve looking at the current threshold for racial disharmony under section 61 of the Human Rights Act, and for inciting hatred under section 131 of the Crimes Act. The current thresholds are high compared to other comparable jurisdictions and relatively few prosecutions have been carried out. Under the incitement provisions, there has only ever been one case.

The Human Rights Commission supports a proposed review of our regulatory and legislative protections against hate speech. There are a number of areas that could be looked at. For example, the current threshold for racial disharmony under section 61 of the Human Rights Act, and for the criminal offence of inciting hatred found in section 131 of the Act. The current thresholds are high compared to other comparable jurisdictions and relatively few prosecutions have been carried out. Under the incitement provisions, there has only ever been one case.

"In other comparative jurisdictions," explains John Hancock of the Human Rights Commission, "there are generally more cases, and more criminal provisions when it comes to hate speech. We just have this one provision, it's a very high threshold, it's never used. Why is that?"

"There does need to be a high threshold," he says, "but does that need to be at the level where you have to both intend and be likely to incite contempt or ill will against the group, or could it be a lower? Could there be, for example, a recklessness standard as we see in other jurisdictions." The Commission does not have a set position on these points currently but stresses that these are the types of questions that need to be asked and considered by our community when looking at whether our current hate speech laws are adequate.

CHANGE UNDERSTANDING & BEHAVIOURS

The Human Rights Commission has also highlighted anomalies about the scope of the characteristics protected under section 61 and 131 of the Human Rights Act. Questions are being asked about whether these should be extended to include religious discrimination and discrimination against people with disabilities or on the basis of sexual orientation or gender. The current provisions are based on the Convention on the Elimination of Racial Discrimination, so they focus on race, to the exclusion of other characteristics which are known to be the target of both discrimination and hatred and which are included in other provisions of the Human rights Act. Eleanor Vermunt, Legal Advisor at the Human Rights Commission, explains that section 61 of the Human Rights Act sets out protected characteristics for the purpose of defining hate speech and at the moment they include only race, nationality and ethnicity. In Canada, she explains, all the grounds in the Human Rights Act, including disability, are protected characteristics for the purpose of defining hate speech.

Some interviewees suggested it was time for a review of all the various pieces of legislation designed to protect citizens against online hatred and abuse, including the relatively recent Harmful Digital Communications Act, and the various institutions mandated to monitor and implement those laws. This review would identify overlaps and gaps in the current regulatory environment – such as the lack of adequate protections for online abuse by and of groups, rather than individuals. It could propose ways to integrate the various mechanisms and provisions in order to make them easier for people to navigate, and ideally look for ways to rely less on individual complaints and reporting as the means to monitor online abuse.

EDUCATION AND BEHAVIOUR CHANGE

Some interviewees thought that increasing education and digital literacy would help in mitigating the negative impacts on our democracy of radicalisation and misinformation through digital platforms.

Joel Colon Dios argues that radicalisation can't be protected through regulation. "This is actually about politics, and having an educated population that can distinguish between what is reliable information." Paul Moon also sees a role for education in reducing vulnerability to radicalisation. "If you don't have an inclination to join a far-right group," he says, "all the advertising in the world won't change your mind."

Moon makes a specific case for an education in history as a means to reduce susceptibility to extremist ideas. "If they had some grounding in history then the curtain would be raised on the origins of a lot of these things. So what you might think is something new and novel, is a reincarnation of something that's been around before." This could act as a counter to the way in which extremist right-wing movements present themselves. "From what I've seen, they put on a very soft, encouraging, engaging front" says Moon, "the rhetoric is all about protecting rights and appeals to nostalgic traditions. But that's just a front for other more nasty things behind it. If you're not aware that it's happened before then you're perfectly open to it happening now." Conversely, he argues, if you are aware that it has happened before, and have seen what that actually looked like and how it played out in history, you are less likely to be open to it.

David Farrar also argues for education as a mitigating factor, and specifically makes a case for solid media literacy training for children. Although he credits young people with already having generally pretty good media skills. “Actually it’s probably the younger ones who can tell their parents, ‘Oh no, no, look it up on snopes.com,’ or ‘No that’s one of those Russian sites.’ They’re probably the ones who actually may have a better idea of what you can and can’t trust.”

Alistair Knott made a case for education for people building digital platforms to better understand the social impacts of their products they build. “We need to have computer scientists educated about the social impacts of the internet products they build. There needs to be ethics and sort of social use insight threaded right through the computer science curriculum.”

One of the arguments that comes up about responses harmful misinformation is whether or not it can be countered with good information. David Farrar represents the view that is can be. “In the end, the solution to bad speech is good speech, though. It’s just trying to make sure you’ve got good information out there to counter that.” He cites the role of the Science Media Centre in trying to counter misinformation and disinformation on topics like vaccination with good quality information.

Other interviewees raised doubts as to whether ‘good information’ was sufficient to counter bad given human cognitive biases. Farrar himself recognised this, referred to a study in the US after Trump won the presidential election. “Trump would say something false, and they’d fact check it, and the fact checks amplified it. They actually concluded it almost helped him.” This is consistent with other research referred to by interviewees who by and large were of the view that supplanting bad information with good was much harder than it might sound.

James Maclaurin says that teaching critical thinking is hard. “I teach global thinking. I have often taught in philosophy departments and I taught for many years.” You can tell people about their cognitive biases, says Maclaurin, but that doesn’t mean that their biases stop working. But there are some ways to get around those biases, he says. “We could agree that I am only going to read about a particular topic when I am with you, because I know you and I disagree about this topic. So, we will balance each other out on this. Or I am going to write down all the reasons why I think this is true. And you are going to write them down too and then we are going to look at them together.” So there are ways to get past the rapid action of our biases, but it takes some work, and isn’t easy to do in the context of digital media as it is currently designed.

For now, and without significant changes in the design of digital media, the kinds of behavioural mechanisms outlined by Maclaurin would probably require individual users to commit to changing their behaviours to regulate their own content and online experiences. This is something that some interviewees recommended as a solution to the problems posed by digital media.

David Farrar says he had to learn to moderate his own behaviour online, and learn that sometimes the best choice he could make was to step away. “I remember at one point someone demanding I answer them online, and I think I was out biking somewhere, and I thought, ‘Am I going to have to pull over the bike?’ Now actually the best thing is to turn the notifications off. Just, turn them off.”

WHAT ROLE FOR NZ?

All of this raises the question: what role should New Zealand play in the wider global efforts to respond to the challenges of digital media? Most of the interviewees for this research were held before the Christchurch mosque attacks, which may have tilted the position of some people in favour of our government playing a more proactive role on the global stage around these issues.

National Party spokesperson on Broadcasting, Communications and Digital Media, Melissa Lee sent through a follow up to her interview after the attacks, noting that the “terrible attack on our country’s Muslim Community took place, aided in part by the malicious use of live-stream technology and digital media platforms” had naturally led to many people in New Zealand becoming “greatly concerned at how information, ideas and content deemed objectionable has been able to circulate in the aftermath”. She notes that the Government is looking at ways to deal with these issues in the digital sector, including on an international scale, and says that “all Members of Parliament will constructively engage” in these issues when they come to be debated.

But even before our country saw with horrifying clarity the dangers of an unregulated digital space, there was a range of opinions amongst those interviewed about the best role for us globally.

Some people argued that New Zealand should follow the lead of bigger, like-minded liberal democracies like the United Kingdom, the European Union and Australia. Others thought that New Zealand could, and should be leading on these issues. And others again saw specific opportunities for New Zealand to provide leadership in certain niches, like preventing mass government surveillance through digital media or protecting indigenous data sovereignty.

One thing that many people agreed on was a sense of some urgency – and urgency which may have increased considerably in the months since many of these interviews took place. As John Edwards put it, “we’ve got some really resounding early warning signals about how this stuff can be used to erode our democratic institutions, and if we don’t sit up and take notice of it, and don’t provide the necessary technical, social, and regulatory responses, we might wake up and find that we’ve missed the opportunity.”

FOLLOWER

One of the main reasons given for taking the path of following the lead of others was New Zealand's size. As Erika Pearson put it, we are tiny. "We are not even a market, we are a sub-market." In terms of negotiating power, Pearson points out, even with the weight of the entire European Union behind it, "it was an uphill push to get the GDPR." So on the one hand you have small national actors "like New Zealand, like the Islands" and then you have these giant global companies. "If Facebook was a country," says Pearson, "I understand it would be approximately the sixth largest by GDP or something ridiculous."

Some participants argue that we need a much better system for making policy on these issues before we can be any kind of global leader. "You've got fairly otherwise smart people just basically giving their opinion with no real information behind it at all. And that's how we make our policy in this space, generally." Either we simply adopt the policy approach taken in another jurisdiction, according to one participant, or we have a relatively flimsy policy discussion which isn't founded in evidence. We need, they say, to build up our capacity as a country to understand and deal with these issues. We need to build up more of an evidence base. "How capable are our public servants at actually tending to this from an evidence-based perspective?" Before we can be leaders in any sense, we need to be equipped to have a solid base for developing policy ourselves.

Thomas Beagle makes a related, but slightly different argument for not taking on any kind of leadership role right now. The behaviours and issues described in this research are emergent, he says, and we're probably going to need to 'muddle on through' for a bit as we see how they emerge. "So possibly there isn't much we can do other than just keep plugging on. Keep on talking to people." In our favour, he argues, is the fact that we are small enough and distant enough that we "may be able to escape the worst of what's happening overseas."

One argument that wasn't made by participants explicitly, but which came up in the context of considering what makes New Zealand different to other countries when it comes to digital media, is the global importance of Chinese and other non-US based social media platforms and messaging apps. Chinese product WeChat, for example, is one of the world's largest standalone mobile apps with over 1 billion monthly active users, with a significant presence here in New Zealand. A question raised by some participants was whether New Zealand was the right country to lead a global effort to address the harms of digital media, and specifically whether the New Zealand government would be able to gain the cooperation of the governments of China, India and Brazil, for example.

LEADER

On the other hand, some interviewees asked why New Zealand should be a 'taker' of policy on these issues. "There is a great opportunity for New Zealand to team up with other like-minded democracies," says Andrew Butler, "and say 'what are we going to do?' [W]e are typically at the cutting edge of technology, so why would we not participate?" If we have a particular perspective, and Butler argues that in New Zealand we do, "we should put it out there." Digital media has brought advantages to New Zealand, says Butler, by reducing the feeling of distance. "It's a really good thing for our community, so we want to make sure that we don't lose the upsides of the new digital economy. We'd be crazy to lose that so there's a whole lot of reasons why we would want to get involved."

Joy Liddicoat says in the past New Zealand's government has not been engaged in the digital rights debates internationally. "When I was in the United Nations working on this between 2011-2014. [New Zealand government representatives] were absent from debates on media freedom. They were absent from debates on human rights defenders, and journalists." This had started to shift with the new Labour-led government, says Liddicoat, and with Claire Curran's interest in digital rights, which was positive. But the challenge is still there for digital rights advocates, says Liddicoat, to find ways to talk about these issues which help people see the relevance to their daily lives, and the issues they face regularly online.

NICHE INFLUENCER

Some interviewees pointed to New Zealand's track record of taking a principled stand on big global issues, giving our nuclear free policy as an example. Richard Macmanus was reminded of that history as he watched the country's response to Stephen Molyneux and Lauren Southern when they visited New Zealand in 2018. "The fact that they were effectively driven out of here without really being able to do their thing makes me feel a bit proud. It reminded me a little of the anti-nuclear stance. I think New Zealand as a country can take stances like that, 'We don't want that kind of propaganda in our country.'" So maybe that is an opportunity for NZ, says Macmanus, "because we've always been known for taking a principled stand about things, maybe that's something we can do."

Researcher Tom Barraclough sees an opportunity for New Zealand to lead on responses to deepfakes and synthetic media. "New Zealand can be an international leader on how legal, technological and policy systems can maximise the benefits and limit the harms of synthetic media," he says. "It can lead by demonstrating a measured and proactive response to examining its existing legal system as well as stimulating access to forensic services to empower citizens to utilise that system."

One area in which New Zealand could show leadership, says Nat Dudley, is in the development of a tech workers' union. "There is actually absolutely movement on that in New Zealand." The problem, she says, is New Zealand technical designers and workers are not the ones making the decisions about the design of the major platforms. "We can influence it at a New Zealand scale." And because New Zealand has comparatively better employment protections than many other places where tech people work, she says, "we already have less of the fear of speaking up. We also have a small enough sector where personal relationships can very easily be brought to bear on these situations. And there are enough senior people in positions in the sector who have more ethical. I don't think we're in a good place but we're in comparatively a better place I think."

On a related topic, Simon Howard points to the leadership seen from the big telecommunication companies like Spark and Vodafone in the wake of the Christchurch mosque attacks. 'We've seen some great leadership from them,' says Howard referring to their decision to block access to 4chan, liveleak, 9gag and 8chan where copies of the video of the attack were circulating. "I think that's what needed to happen," says Howard, "they wore some slack for it, but Simon Moutter's been awesome on Twitter. He's just been giving regular updates, saying it's only temporary, explaining 'this is what we're protecting people from seeing'."

Thomas Beagle points to another area in which, he says, New Zealand is currently doing relatively well. “I look around the world and I feel that we’re doing better than most.” In terms of government surveillance, which is an area of particular interest to Beagle as a defender of civil liberties, “we’ve been making steps both forward and backwards, but I think we’re still holding back from total government surveillance. What we’re doing in the Pacific is a worry, but looking at what we’re doing to ourselves here we are ahead of many.”

Finally, but perhaps most importantly, some interviewees argued that there was a role for New Zealand to play as a leader on indigenous data sovereignty and issues relating to Māori digital issues. This would first require us to address the significant gaps in our own protection of indigenous rights online. Karaitiana Taiuru has outlined a range of areas where improvement is needed and has suggested some solutions. One of the most critical issues is the need to protect indigenous data sovereignty, allowing Māori ownership and control of Māori data.

Taiuru asks whether Māori data and data relevant to Māori is being adequately protected as a taonga in compliance with the Treaty. And whether Māori data sovereignty is being enabled. One possible solution to this is an approach similar to that taken on the Solid website (<https://solid.mit.edu/>). Another area where better protection is needed is in online abuse, Taiuru says that Netsafe are not currently equipped to deal with tikanga Māori issues online or even deal appropriately with Māori victims of online abuse. There is a need for cultural and ethical considerations of te ao Māori in developing digital security policy. Further discussion and protections are also needed around the digital colonisation of Māori culture and values. Overall, says Taiuru, appropriate consultation and engagement is needed by Netsafe and other organisations with a diverse range of Māori around digital issues affecting Māori including cyber safety, online abuse, data management.

APPENDIX 5: ONLINE HATE AND OFFLINE HARM

> >
> > >
> > > > > > > > > >
> > > > > > > > > > > > >
> > > > > > > > > > > > >
> > > > > > > > > > > > >
> > > O N L I N E > > > >
> > > H A T E > > > > > > >
> > > > > A N D > > > >
> > O F F L I N E > > > >
> > H A R M > > > > > > >
> > > > > > > > > > > > >
> > > > > > > > > > > > >
> > > > > > > > > > > > >
> >
> > > > >
> >

ONLINE HATE AND OFFLINE HARM

This research was funded by InternetNZ, and prepared by Jess Berentson-Shaw & Marianne Elliott of The Workshop.

“The web is under threat — but the web we want is not out of reach. It’s up to us to overcome these threats and ensure the web remains an open platform that is truly a force of good for everyone.”

- The Case for the Web

CONTENTS

BACKGROUND	231
OVERVIEW	232
1. HATEFUL AND RACIST DISCOURSE AND HATEFUL CRIMES OR TERRORISM	233
2. FREEDOM OF PEOPLE TO ACCESS GOOD INFORMATION	236
3. THE SPREAD OF RACIST IDEAS	238
4. IMPACTS ON PEOPLE	240
4.1 DIRECT BENEFITS FROM THE INTERNET	240
4.2 DIRECT HARM FROM THE INTERNET	241
4.3 INDIRECT HARM FROM RACIST DISCOURSE ON POLICY TO ADDRESS THE PROBLEM	242
5. TE TIRITI O WAITANGI	243
CONCLUSIONS	244

BACKGROUND

The shock and horror felt by many New Zealanders at the Christchurch mosque massacres has reignited a debate in our country about the role of the internet in fostering and spreading hatred generally, and extremist violence specifically. Much of the public commentary on this issue is currently framed as a discussion about content moderation, with little reference to or understanding of the larger challenges facing the open web.

This presents an important opportunity for New Zealand to contribute to a broader, more well-informed and evidence-based conversation about the relationship between online hatred and real-life harm.

In its report, 'The Case for the Web' the World Wide Web Foundation, founded by Sir Tim Berners-Lee, outline the major threats currently facing the open internet. Those threats include:

- a massively disproportionate power over online spaces in the hands of a few, huge companies.
- online decisions with serious real-life consequences are increasingly being made by algorithms and machines, replicating biases and reinforcing inequalities found offline.
- the web is not "for everyone" and most of the unconnected are marginalised populations in low- and middle-income countries, and most are women.
- a decline in the ability of anyone, anywhere to create and share content, ideas and information freely, driven by exposure to online harm including through misuse of data, malinformation and online hatred and abuse.

This paper focuses largely on the fourth of these main threats, the role that online harm plays both in undermining the open nature of the web and in creating offline harm. However, inevitably, a consideration of any one of these threats overlaps with the others, since the ways in which online hate has been spread is linked to both the role of algorithms in online decisions and the dominant role a handful of companies, including Google and Facebook, increasingly play in most people's experience of the web.

OVERVIEW

This paper looks at the evidence as to the role online hate plays in spreading racist thinking and actions offline, including hate and terror crimes. It explores the evidence for whether harmful action online has increased both harmful ideas and actual harm against minority ethnic and racial groups.

Specifically we cover

- Hateful discourse on the internet and acts of violence,
- Freedom of all people to access and use quality information (including that related to issues of race and ethnicity),
- Spread of racist ideas and ways of thinking across populations and,
- Impacts on people from minority ethnic and religious groups in everyday life.

We start with a discussion on direct links between hateful discourse and group membership on the internet and acts of extremist violence. However, we emphasize that this is a very high burden of proof: most people do not commit acts of extreme violence making it harder to measure and research. The research evidence shows a complex systems effect at play in the relationship between online hate and offline harm, which requires we look beyond direct surface links between hate on the internet and acts of extreme violence when exploring the role and responsibilities of the open internet.

1. HATEFUL AND RACIST DISCOURSE AND HATEFUL CRIMES OR TERROSISM

Empirical literature shows extremism has multiple causes like most human behaviour. The question is: does the internet play a role? The answer is certainly yes according to those who research these connections¹. The mainstream social discourse, of which a large component now happens online, plays a fundamental part in acts of hate, violence and terrorism.² Mainstream discourse on the internet acts to network hate, not disinfect it via sunlight (as we discuss later).

The question is not if the internet plays a role in violent extremism, but how large a role it plays specifically in the contemporary radicalisation process³. Research from 2017 showed “while self-radicalisation is possible through the medium of the internet, physical contact played a significant role in the case of those interviewed. This evidence, while based on a small number of cases, also shows that while the internet facilitates the radicalisation of individuals it is not the sole driver of the process.”

Chan et al studied the link between 14,000 internet sites containing hate related content and official racial hate crimes and found high correlations, most evident in areas with higher racism and segregation. They also found a link between online hate sites and those racially motivated crimes by lone wolf perpetrators⁴. The Southern Poverty Law Centre noted in 2014 that active users of the website Stormfront.org were allegedly responsible or the murders of nearly 100 people in the previous 5 years.⁵

Raphael Cohen-Almagori has made extensive case studies of acts of race hate and violence and concludes “there is sufficient evidence to suggest that speech [on the internet] can and does inspire crime” and “Bigots, inspired by what they have read online, went on to inflict violence on their targets.” He includes analysis of the connection

1 Raphael Cohen-Almagori. “Taking North American White Supremacist Groups Seriously: The Scope and the Challenge of Hate Speech on the Internet”. International Journal of Crime, Justice, and Social Democracy (2018).

2 See the safehousealliance toolkit <https://www.safehousealliance.org/wp-content/uploads/2012/10/Tools-for-Liberation-Packet-SPAN.pdf>, based on 2005 research, by Paula Rothenberg. <http://fabletext.info/paula-s-rothenberg-race-class-and-gender-in-the-united-states-an-integrated-study-eighth-edition-fable-text.pdf>

3 Conway, Maura. “Determining the role of the internet in violent extremism and terrorism: Six suggestions for progressing research.” Studies in Conflict & Terrorism 40.1 (2017): 77-98.

4 Chan J, Ghose A and Seamans R (2016) The Internet and racial hate crime: Offline spillovers from online access. MIS Quarterly 40(2): 381-403.

5 Beirich H (2013) Hate across the waters: The role of American extremists in fostering an international white consciousness. In Wodak R, Mral B and KhosraviNik B (eds) Right Wing Populism in Europe. Politics and Discourse: 89-104 London: Bloomsbury.

between hate groups and biochemical warfare in particular.⁶

One study looked at a sample of fifty-one Canada extremists to examine the role of social media both in the process of radicalization, and how extremists use social media after they become radicalized.⁷ The results confirm that social media played a role either during or after the radicalization process of the majority of the sample and converts are more vulnerable to online radicalization than non-converts.

Nikita Malik, Director of the Centre on Radicalisation and Terrorism (CRT) at the Henry Jackson Society, has reported on recent (unpublished) research into which social media platforms were being utilized to spread hatred.⁸ She found that people who were convicted for spreading material and planning hate attacks in the Far Right tended to use only two platforms to do this – Twitter and Facebook. By comparison, her previous research found that Islamist related content was spread on a variety of platforms including the Darknet and encrypted messaging services.⁹

In general, what the observational research shows is that case for the link between hate speech on the Internet and hate crimes has been well made, however more research is needed to understand the details.

We can also look to intervention research for information that observational studies may not give. Evidence shows that intervening in the spread of racist and hateful ideas and ways of thinking at multiple levels (including in public discourse which includes the internet) works to prevent terrorist acts. Figure 2 below outlines the macro, meso and micro level interventions that have evidence of impact on countering extremism (translated from German).¹⁰

6 Cohen-Almagor, Raphael. "Balancing Freedom of Expression and Social Responsibility on the Internet." *Philosophia* 45.3 (2017): 973-985.

7 Mehmet F. Bastug, Aziz Douai, Davut Akca (2017), Exploring the "Demand Side" of Online Radicalization: Evidence from the Canadian Context. *Studies in Conflict & Terrorism*.

8 Malik, N. Unpublished research reported in Forbes (2018), The Fight Against Terrorism Online: Here's The Verdict. Retrieved from <https://www.forbes.com/sites/nikitamalik/2018/09/20/the-fight-against-terrorism-online-heres-the-verdict>

9 Malik, N. (2018) Terror in the Dark: How Terrorists use Encryption, the Darknet and Cryptocurrencies Retrieved from <http://henryjacksonsociety.org/wp-content/uploads/2018/04/Terror-in-the-Dark.pdf>

10 Daniel Köhler & German Institute on Radicalization and De-Radicalization Studies (GIRDS). Structural Quality Standards for work to intervene with and counter violent Extremism. A handbook for practitioners, state coordination units and civil society programme implementers in Germany. <http://girds.org/publications>

Figure 2:

THE COUNTER-TERRORISM NETWORK

	MACROSOCIAL	MESOSOCIAL	MIKROSOZIAL
PREVENTION (GENERAL, SPECIFIC)	e.g. education, research, youth work, social work	e.g. local crime prevention, federal state action plans (LAPs) etc.	e.g. workshops in schools with former extremists
REPRESSION	Legislative, executive, nationwide security architecture	e.g. banning orders, neighbourhood officers	e.g. banning orders, neighbourhood officers
INTERVENTION	Counter-narratives	e.g. family support	e.g. individual exit programmes

The following sections on the role of the internet in blocking access to good information, mainstreaming racist and white supremacist ideas from hate groups, and influencing public discourse and support for policies that address bigotry, racism and their impacts, highlight the importance of the role of the internet on hateful discourse and actions.

We emphasize the need to examine closely arguments that require a absolute 'proof' for the internet being a sole cause of an act of terrorism in order to take action. It does not reflect best scientific practice in an area as complex as this. And it may reflect the effects of the laundering and spread of harmful ideas and ideology into the mainstream public discourse, which plays a key role in undermining support for policies and practices that could limit bigotry and racism, and their harmful impacts, as we discuss later.

2. FREEDOM OF PEOPLE TO ACCESS GOOD INFORMATION

The evidence is clear that the internet has democratised the availability of information. However, the benefits derived from greater availability of information is countered by an infrastructure that make it increasingly difficult for people to locate and recognise good quality information.

Specifically, research shows that both the technical effect of the attention economy (advertising revenue linked to attention) and social psychological processes work to direct people using search engines and social media towards disinformation, misinformation, and material that confirms their existing incorrect or conspiratorial beliefs.¹ Echo chamber effects on Twitter, for example, work to restrict people's exposure to ideologically challenging discourse in particular.²

False information, particularly that of a polarized and political nature, evokes strong emotions, and hold people's attention. Taking advantage of basic human social and psychological processes, leading internet platforms have developed their core technology to capitalize on this behaviour, as spreading such information drives advertising revenue.³ Cambridge Analytica shows that targeted political advertising is a highly profitable business.⁴ Researchers found that false political news and viral types, notably terrorism, diffuses faster, deeper and more broadly than the truth on the internet.⁵

Even when people are made aware of misinformation and false information, their individual ability to identify and counter this is limited and often ineffectual due to the core technologies that spread misinformation. Tripodi undertook a study of conservative Christians in the US, who sought to critically interrogate media, news and other information, by fact checking source material against multiple internet sources.

1 Ghosh, D., & Scott, B. (2018). #digitaldeceit: The Technologies Behind Precision Propaganda on the Internet. Retrieved from <https://www.newamerica.org/public-interest-technology/policy-papers/digitaldeceit/>, Srnicek, N. (2017). Platform capitalism: Cambridge, UK: Polity Press.

2 Guo, L., Rohde, J. A., & Wu, H. D. (2018). Who is responsible for Twitter's echo chamber problem? Evidence from 2016 U.S. election networks. *Information, Communication & Society*, 1-18., Himelboim, I., McCreery, S., & Smith, M. (2013). Birds of a Feather Tweet Together: Integrating Network and Content Analyses to Examine Cross-Ideology Exposure on Twitter. *Journal of Computer-Mediated Communication*, 18(2), 40-60

3 Ghosh, D., & Scott, B. (2018).

4 Isaak, J., & Hanna, M. J. (2018). User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection. *Computer*(8), 56. doi:10.1109/MC.2018.3191268

5 Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146-1151.

The author found that “fact checking” returned further false information because 1) search terms used by people are biased, 2) search engines are returning results that are far-right or alt-right to mainline conservatives without them knowing, and 3) bad faith actors are exploiting the intellectual exploration of conservatives.⁶

Alongside the democratisation of information there has been no commensurate democratisation in the structures, tools, or an individual’s ability to accurately identify (or assess) quality information.

⁶ Tripodi F. Searching for Alternative Facts. Analysing Scriptural Inference in Conservative NewsPractices. 2018. Data & Society. Accessed May 2018 <https://datasociety.net/output/searching-for-alternative-facts/>

3. THE SPREAD OF RACIST IDEAS

Research shows that the loose and networked effects of the internet have acted to “launder” racist, and notably white supremacist, information into the mainstream. Dominant online structures do not expose these ideas to the sunlight of an open space, rather research shows they sanitise them and hide them in plain sight, so people cannot recognise them for what they are.¹ Figure 1 below describes the process of “information laundering” of hate speech.

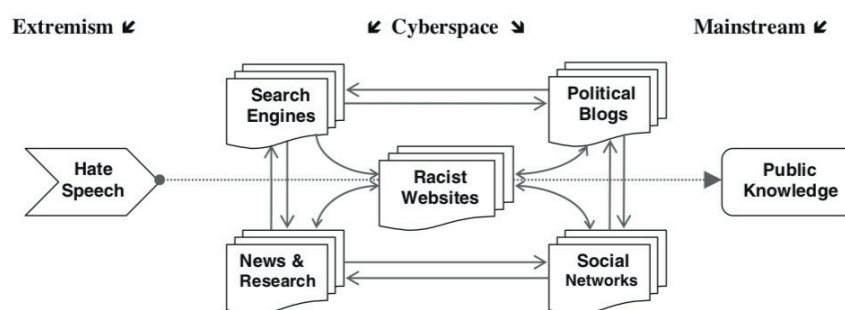


Figure 1 Model of information laundering.

Communication Theory 22 (2012) 427–448 © 2012 International Communication Association

435

“Through the Internet, hate groups are entering into mainstream culture by attaining legitimacy from the established media currencies of the cyber community, primarily search engines and interlinking social networks. These conventional pathways can unwittingly lead an online information seeker to extremist content that....has already been designed for them to appear as educational, political, scientific, and even spiritual in nature. More importantly, however, this network can also allow content from these websites to travel outside of their domain, to merge with mainstream spaces like YouTube, Facebook, political blogs, and even occasionally news cycles.” Klein, 2012

For example multimedia convergence adds legitimacy to the white power websites, especially for young people who are familiar with and trusting of specific internet sources and brands. A study of 26 white supremacist websites in 2010 showed that YouTube videos appeared frequently, for example the creativityalliance YouTube channel.² Cognitive research shows that people have a poor memory for source material,

¹ Klein, Adam. “Slipping racism into the mainstream: A theory of information laundering.” Communication Theory 22.4 (2012): 427–448.

² Klein, A. G. (2010). A space for hate: The White Power movement’s adaptation into cyberspace. Duluth, MN: Litwin Books, LLC.

often misattributing bad information to trusted sources.³ When those sources converge with malicious sources the problem is exacerbated.

The emergence of bias in the media is another effect of information washing. A meta analysis in 2017 found that “that Muslims tend to be negatively framed, while Islam is dominantly portrayed as a violent religion”.⁴

The case of VDARE.com shows how information laundering works in reality. VDARE.com emerged in 1999 as an anti-immigrant website in the US, and by 2003 was added by the Southern Poverty Law Center as a hate group. It both attracted pro-white communities and published anti-Hispanic and anti-Semitic essays. These online essays take the form of “scholarly texts” and become certified in the public domain, research that then becomes cited in media by commentators, influencing the public discourse and thinking. By 2012 VDARE had reached a mainstream audience, Its founder invited to speak at Conservative Political Action Conference (CPAC).

“In this objective, the Internet becomes the key to activating the process of information laundering, turning hateful rhetoric into public knowledge.” Klein, 2012

3 Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N., & Cook, J. Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, (2012) 13(3), 106-131.

4 Ahmed, Saifuddin, and Jörg Matthes. “Media representation of Muslims and Islam from 2000 to 2015: A meta-analysis.” *International Communication Gazette* 79.3 (2017): 219-244.

4. IMPACTS ON PEOPLE

Online structures have the potential to both facilitate positive experiences for people of minority ethnic and racial groups and drive harm against them. The balance is currently tipped far towards harm.

4.1 DIRECT BENEFITS FROM THE INTERNET

The examination of several case-studies in the Australian context shows digital media can facilitate the formation of both ‘ad-hoc’ and longer-term, group-based online communities focused on fighting racism, for example IndigenousX.¹ Social media platforms, such as Twitter have been found to play a key role in the organisation of protest movements, like that of Arab Spring, Los Indignados and Occupy movements.² Social media may act to reduce political inequality, through encouraging political engagement.

The Workshop’s own research in New Zealand shows social media is used by minority ethnic groups to access politicians and political action (unpublished survey research).

However, while evidence suggests the internet has given minority groups greater access to similar voices, collective action and political actors, it has also exposed them to more hate and bigotry. In 2017, one in ten New Zealanders experienced hate speech online and three in ten encountered hateful content.³ One in three New Zealand women experienced 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.⁴ In 2018, one in three Māori (32%), and one in five Asian (22%) and Pacific (21%) people experienced racial abuse and harassment online.^{5 6}

1 Jakubowicz, A., Dunn, K., Mason, G., Paradies, Y., Bliuc, A.-M., Bahfen, N., Connolly, K. (2017). *Cyber Racism and Community Resilience: Strategies for Combating Online Race Hate*: Springer.

2 Bennett, W. L., Segerberg, A., & Walker, S. (2014). Organization in the crowd: peer production in large-scale networked protests. *Information, Communication & Society*, 17(2), 232-260.

3 Netsafe, (2018) *Online Hate Speech A survey on personal experiences and exposure among adult New Zealanders*.

4 Amnesty International New Zealand (2018), *Amnesty reveals alarming impact of online abuse against women*

5 ActionStation (2019), *The People’s Report on Online Hate, Harassment and Abuse*.

6 Netsafe has published two additional reports which reflect the findings from the Amnesty and ActionStation reports, and in addition include people in the 14-17 age group, and provide a measure of the distress and harm caused. They are: *Netsafe (2018) New Zealand Teens and Digital Harm: Statistical Insights into Experiences, Impact and Response* and *Netsafe (2017) Harmful Digital Communications in New Zealand: Annual Population Survey 2017*. A Netsafe third report included findings about the experience of non-heterosexual adults (e.g., young adults (aged 16-29) and those who identify as non-heterosexual were more likely to be targeted by Image-Based Sexual abuse). *Netsafe (2019) New Zealand Teens and Digital Harm: Statistical Insights into Experiences, Impact and Response*.

4.2 DIRECT HARM FROM THE INTERNET

Hate speech pre-dates the internet. However, alongside the rise of digital technology, there has been a 66% increase in hate groups in the US since 2000 (602 groups to 1002).⁷ In 2011 there were over 14,000 documented hate sites, blogs, and social networks operating across the web.⁸

The structures and features of dominant online tools and platforms, including capitalization of people's attention, the design of search engines and social media algorithms, networking effects across websites, media and social media, allow for increasingly coordinated and targeted hate towards minority-ethnic groups. Research also found the use of Encryption, the Darknet and Cryptocurrencies to engage in activities of hate and extremism.⁹

Racialised hate speech (otherwise known as cyber racism) is specifically targeted towards ethnic minority groups, and has become increasingly coordinated in recent years, through the rise of the "alt-right". It has become a global phenomenon, affecting "refugees and ethnic minorities in Europe, Muslim Blacks and Jews in the United States, Indigenous Australians" and others. The impact on affected groups is to retreat to safe locations to find solidarity, instead of engaging in public discourse and institutions, where thinking may be changed.¹⁰ Spirals of silence have been found in large surveys, where majority groups act to suppression of minority groups dissenting views and force withdrawal into safe spaces.¹¹

Two social phenomenon should be mentioned here - pluralistic ignorance. Where the amplifying effects of the internet have been found to push those with more moderate views to polarised extremes due to a belief extreme views are more common than they are. And the false consensus effect, where by those with hateful and racist views believe their views to be more widely shared.

Anne Pedersen has carried out extensive research into the beliefs and attitudes about minority groups in Australia.¹² In one of her studies people were asked about their attitudes to Aboriginal Australians and asylum seekers. They were also asked to estimate what percentage of people they believed held the same attitudes as them. Whether an attitude was positive or negative, people tended to overestimate the community support for their own beliefs. However, those who held the most fringe and negative beliefs about these two groups of people, vastly overestimated the support for their beliefs in the community compared to those with positive attitudes. Those with positive beliefs estimated 46% of the community shared their attitudes, those with negative beliefs

7 U.S. Hate Groups Top 1000 (2011). Retrieved from <http://www.splcenter.org/get-informed/news/us-hate-groups-top-1000>

8 Simon Wiesenthal Center. (2011). 2011 digital terrorism & hate report launched at Museum of Tolerance New York. Retrieved from <http://www.wiesenthal.com/site/apps/nlnet/content2.aspx?c=lsKWLBpJLnF&b=6478433&ct=9141065>

9 <http://henryjacksonsociety.org/wp-content/uploads/2018/04/Terror-in-the-Dark.pdf>

10 Jakubowicz, A., et al. (2017). *Cyber Racism and Community Resilience: Strategies for Combating Online Race Hate*: Springer.

11 Chen, H.-T. (2018). Spiral of silence on social media and the moderating role of disagreement and publicness in the network: Analyzing expressive and withdrawal behaviors. *New media & society*, 20(10), 3917-3936

12 Pedersen, Anne, Brian Griffiths, and Susan E. Watt. "Attitudes toward Out groups and the perception of consensus: All feet do not wear one shoe." *Journal of Community & Applied Social Psychology* 18, no. 6 (2008): 543-557.

estimated 71% of the wider community shared their beliefs. The actual figure is 1.9% for the most extreme negative views. The more negative an attitude towards Aboriginal Australians or asylum seekers, the greater the estimate of community consensus became.

4.3 INDIRECT HARM FROM RACIST DISCOURSE ON POLICY TO ADDRESS THE PROBLEM

Washing information into the mainstream helps build and maintain a public environment less likely to support policies that protect the freedoms and rights and well-being of minority groups. It furthers shallow understanding of minorities lives and experiences, focussing instead on explanations that include individual and moral failings. This causes genuine harm - policies that fail to address long term negative outcomes, and reinforce them. A large scale analysis of empirical research in health Williams and Mohammed show how different aspects of racism affect wellbeing.¹³

First, Williams and Mohammed found evidence that racism at an institutional level leads to policies and actions that limit Black Americans' access to important resources and opportunities; for example, by restricting access to certain neighbourhoods, education, employment and other community resources (also observed in New Zealand research¹⁴). Second, racism embedded in cultural narratives (such as internet, media and popular culture) shapes negative emotions about Black people, and leads to stereotypes and prejudice that damage people's wellbeing. Finally,

they found a large body of evidence showing that "experiences of racial discrimination are an important type of psychosocial stressor that can lead to adverse changes in health status and altered behavioural patterns that increase health risks."

The authors draw our attention to the role that the absence of positive feelings towards stigmatised groups has in shaping policy preferences of wider society. The cultural narratives create a social and political environment hostile to policies that attempt to change the social and economic conditions that drive harm. Altered behaviours and health status that stem from wider social conditions are interpreted by others through the narratives of discrimination. These narratives posit that it is inferiority, biology and/or individual weakness that determine poorer outcomes, not social, economic, cultural or environmental conditions. This interpretation prevents support for action, such as the introduction of policies and regulations, to change those conditions.

¹³ Williams, D. R., & Mohammed, S. A. (2013). Racism and health I: Pathways and scientific evidence. *American Behavioral Scientist*, 57(8), pp. 1152-1173.

¹⁴ Harris, Ricci, et al. "Racism and health: The relationship between experience of racial discrimination and health in New Zealand." *Social science & medicine* 63.6 (2006): 1428-1441.

TE TIRITI O WAITANGI

While out of the scope of this paper, we recommend that an assessment be made of the impact of online hatred and abuse on the rights and protections set out in Te Tiriti o Waitangi, as a foundational document of New Zealand.

CONCLUSIONS

The evidence for the connection between online hate and ‘real-life’ harm is clear: there is a connection, albeit a more complex one than some have hypothesised. And although we need more research to better understand the details of how that connection works, sufficient evidence already exists to make the case for action to, as Sir Tim Berners-Lee has put it ‘save the web’.

As The Case for the Web report concludes, ‘The web is under threat - but the web we want is not out of reach. It’s up to us to overcome these threats and ensure the web remains an open platform that is truly a force of good for everyone.’

Specifically in relation to online hatred and abuse, The Case for the Web report recommends that governments and companies must enact policies and enforce regulations to protect the right to safety alongside the right to freedom of speech.

In New Zealand, the public conversation about protecting the right to safety alongside the right to freedom of speech has too often been overly simplistic and has ignored the growing evidence base on the clear yet complex role that online hatred plays in creating, spreading and sustaining ‘real life’ harm. ‘Real life harm’ is too often understood only in the form of acts of violent extremism. Obviously such acts are a very serious concern, and have a particular relevance to public discourse in New Zealand at the moment.

However, research shows that online hatred and bigotry also do real-life harm in other, more complex and interrelated ways including - amongst others:

- through the ‘laundering’ effect of hiding extreme views plain site in mainstream public discourse,
- through the direct psychological harm of encountering hatred online, and
- through the influence of racist and bigoted discourse on support for policies to protect the freedoms, rights and well-being of minority groups.

The current debate about the role of harmful online content in perpetrating real-life harm provides a critical opportunity to refocus the debate away from a false dichotomy between the right to safety and freedom from discrimination, on the one hand, and the the right to freedom of expression on the other.

New Zealand is currently well-placed to contribute to a broader, more well-informed and evidence-based conversation about the relationship between online hatred and real-life harm. There is an important role for New Zealand to play in reclaiming the values of the open web, and reasserting the critical components of an open internet.

If the open web is to be saved, it’s salvation will lie in a broader public understanding of the critical role of open and diverse internet governance, open architecture and open technology rather than the current focus on a narrowly defined, and widely misunderstood and misrepresented, concept of freedom of expression.

Broadening this conversation, and ensuring that it is informed by the best available evidence, is critical work.

