ICT for peace foundation

POLICY BRIEF

THE NATURE OF DISINFORMATION AND NURTURE OF DEMOCRATIC RESPONSES

REFLECTIONS, RESEARCH AND WRITING ON AOTEAROA NEW ZEALAND'S CHRISTCHURCH MASSACRE

Sanjana Hattotuwa (Author)

Daniel Stauffacher (Editor)

GENEVA 2021

ICT4Peace Foundation

ICT4Peace Publishing, Geneva. August 2021
Copies available from www.ict4peace.org

THE NATURE OF DISINFORMATION AND NURTURE OF DEMOCRATIC RESPONSES

REFLECTIONS, RESEARCH AND WRITING ON AOTEAROA NEW ZEALAND'S CHRISTCHURCH MASSACRE

Sanjana Hattotuwa (Author)

Daniel Stauffacher (Editor)

ICT for peace foundation

CONTENTS

FOREWORD

ICT4Peace is delighted to publish this compilation of a series of short papers, articles, op-eds by Sanjana Hattotuwa, ICT4Peace Special Advisor, in the aftermath of the horrendous March 2019 Christchurch massacre in New Zealand and the April 2019 Sri Lanka Easter Bombings. Since March 2018, Hattotuwa is pursuing doctoral research on social media and violence at the University of Otago, New Zealand.

Covering a period of three-years, Hattotuwa's contributions in New Zealand's domestic media, international media and other public fora including lectures in Europe critically examine a spectrum of issues brought to light by the 2019's violence and terrorism. Reducing harm online, how offline context influences online content, the role of disinformation, the reach of misinformation, addressing content inciting violence, the opportunities and challenges posed by AI in the reduction of online harms, domestic media regulation, and human rights considerations in platform governance are some of the issues covered in this collection, featuring op-eds, essays, policy briefs, presentations, lectures and podcasts.

The ICT4Peace Foundation was also honoured to be invited by New Zealand's Prime Minister Jacinda Ardern and the French President Emmanuel Macron to contribute to and participate[1] in the Christchurch Call to Action Summit, which took place on 15 May 2019 in Paris, France, two months after the two attacks.

In its contribution to the Christchurch Call to Action Summit, ICT4Peace was able to build on its close and fruitful cooperation with UN Counter Terrorism Executive Directorate (CTED) from 2015 to 2019.[2]

1 https://ict4peace.org/activities/ict4peace-was-invited-by-rt-hon-jacinda-ardern-to-discuss-christchurch-call-to-action-to-eliminate-terrorist-and-violent-extremist-content-online/

2 Joint CTED-ICT4Peace activities included ICT4Peace moderating in December 2015 the first ever technical panel at the level of the UN Security Council on collaboration between the Public and Private Sector to promote safety and counter messaging on the Internet, to prevent the recruitment of terrorists and incite terrorist acts, while respecting human rights and fundamental freedoms, which led - inter alia - to the launching by UNCTED and ICT4Peace of the Tech against Terrorism Project, the publishing of the groundbreaking report "Private Sector Engagement in Responding to the Use of the Internet and ICT for Terrorist Purposes Strengthening Dialogue and Building Trust" and the co-hosting in August 2017 of the U.S. launch of the Global Internet Forum to Counter Terrorism (GIFCT) at Swissnex in San Francisco

Contemporary debates around online harms and content inciting hate reflect the output and work of the Foundation over many years, including with the United Nations, partners and various bilateral and multilateral platforms in many countries. Through Hattotuwa's writing, alongside the wealth of related material produced by him and colleagues on the Foundation's website, ICT4Peace continues to reflect critically and add to significant developments in New Zealand, Switzerland, Europe and beyond.

The Christchurch massacre, resulting in the Christchurch Call as a global platform, I have no doubt will help the world move towards rights-respecting processes to challenge the growth of online harms that increasingly impact all of us, and indeed, now lie at the heart of democracy. Hattotuwa is no stranger to violence, offline and online. His research, writing and insights, shared freely and to wide critical acclaim, are deeply valued at the Foundation.

This compilation reflects the core values of ICT4Peace, to explore and champion the use of ICTs and new media for peaceful purposes. We hope you find it as useful as those who have benefitted from this output since 2019.

Daniel Stauffacher
Founder and President
ICT4Peace Foundation, Geneva

OP-EDS

NZ's response can lead the way

OTAGO DAILY TIMES, 20 MARCH 2019[3]

Should New Zealand's response to the Christchurch terror attacks be primarily anchored to national security? asks Sanjana Hattotuwa.

Do you love New Zealand? asked the extremely inebriated young white man and his companion as they suddenly blocked my path at the Octagon, a few weeks after I arrived in Dunedin.

Of all the people on the footpath at the time, I noticed they only followed me for a while. Not knowing quite how to respond, I said affably that I liked what I had seen so far. Entirely uninterested in my answer, coming closer and with their bodies and fingers arching towards me, they said that they were willing to die for New Zealand and that I needed to know this.

Recalling a poet from Sri Lanka who in verse noted that it was far better to live for one's country, I decided just to smile somewhat incredulously. Satisfied that whatever point they had wanted to make had got through to me, they left and lunged into a wine and beer shop.

Claims of New Zealand's innocence lost after Friday's attack in Christchurch need to be tempered with stories that abound around how those who are perceived to be different from or somehow not Kiwi are subject to, every day, the language and looks of condescension, incomprehension and suspicion.

New Zealand is isolated by geography but despite popular belief isn't as exceptional to be immune from ingrained prejudice and latent racism. Some may argue it is an insensitive or inopportune moment to raise more deep-rooted issues, when the more urgent need is to respond to an episode of wanton violence.

An argument can be made to pursue both, recognising that longer-term policymaking requires the unearthing of deep-seated anxieties.

3 https://www.odt.co.nz/opinion/nzs-response-can-lead-way

Fascism thrives on societal insecurities. A document uploaded to the internet by the killer - a self-proclaimed fascist - is instructive reading on this score. First, the language is simple and clear, even if and indeed, particularly because, the logic is so twisted.

The entirely subjective and strategically selective are presented as indubitable fact and authoritative history. Though the document is anchored to right-wing extremism, what's remarkable is how much of it resonates with the anti-Muslim rhetoric spewed by extreme Sinhala-Buddhist nationalist monks in Sri Lanka, and their equivalent in Myanmar. The targets of hate are the same.

Likewise, in the absence of meaningful interaction between diverse groups, faiths, genders or identities, clusters of the like-minded form online, almost immediately putting up high-barriers to inflows of opinion, information and perspectives that contest or question widely held assumptions. Over time, the illusion of diversity based on only the smallest of divergence supplants a more open discussion that embraces radically different ideas.

Author Eli Pariser warned us about this many years ago, noting how algorithmically, our social media accounts feed us what we want to see, instead of what we need to engage with. It is online and by careful design that Friday's fascist found his most receptive audience. As Washington Post journalist Drew Harwell noted: ''The New Zealand massacre was live-streamed on Facebook, announced on 8chan, reposted on YouTube, commentated about on Reddit, and mirrored around the world before the tech companies could even react''.

Policymakers who may not even recognise some of the platforms here have a steep learning curve ahead. New Zealand authorities must now pivot an existing intelligence apparatus geared to hone in on the projected threat of Islamic radicalisation, to more adroitly pick up signals around the very real presence and rapid spread of white supremacist ideology.

Which begs the question: should the response to Friday be primarily one that is anchored to national security?

Coming from Sri Lanka, I sincerely hope not. In my country, legislation purportedly drafted to prevent counter-terrorism has resulted in a convenient framework for successive governments that condones extrajudicial torture and the rampant abuse of human rights, for decades.

The slow erosion of civic rights begins, invisibly, with the emotional appeals to

protect all citizens or certain groups from violence. The necessary balance between proportional responses to new and increasing threats and the protection of civil liberties has escaped Sri Lanka, where more parochial and communal interests have held sway. New Zealand's story, in the months and years to come, must not be this.

Flagged and framed in my social media accounts since Friday is the contrast between a moral and political leadership so visibly present here, yet markedly absent in other countries after a cataclysmic event of this nature, including my own.

What is remarkable since Friday, and reassuring, is the language employed by, and actions of, this country's political leadership. Faced with an unprecedented loss of life, all official responses - without exception - were anchored to denouncing extremism and fringe lunacy, not communities and faiths present in, or part of, New Zealand.

Ironically, it may not even be recognised as exceptional by those living here, but it is precisely that for those of us who are more used to, tired of and frustrated with politicians who are in effect as racist as the terrorists and terrorism they seek to denounce.

Though profoundly distressed by the events of Friday, I am hopeful that the tragedy will result in local and national conversations which lead to, through policy and practice, social, political and cultural templates for other countries to emulate in responding to, and preventing, terrorism.

The encounter at the Octagon fresh into my sojourn in Dunedin was not the only time I have been subject to wary looks and violent language. It is worse for others, identified as belonging to a faith or community that is feared more.

The pain of acknowledging this is - more than or at least alongside revisions to legislation around gun ownership - a necessary step towards a country that may never be able to prevent terrorism, but always sees it as entirely alien to its core values, beliefs and principles rooted in decency, dignity, diversity and democracy.

Terrorists know they have the upper hand on social media

STUFF, 20 MARCH 2019[4]

Coming out of a long meeting, the first I heard of the violence in Christchurch was from those in Sri Lanka who had got breaking news alerts. I was both very disturbed and extremely intrigued.

Terrorism as popular theatre or spectacle is not new, and some academics would argue is a central aim of terrorists, who want their acts recorded and relayed, not redacted or restrained.

The use of social media to promote and incite hate, violence and prejudice is also not new. From ISIS to politicians elected into office through populist, prejudiced campaigns, social media is foundational in contemporary terrorist recruitment and political propaganda.

What events in Christchurch last Friday brought to light was something entirely different, new and very unlikely to be resolved easily or quickly. The killer's intentional use of the internet will have far-reaching implications, requiring significant, urgent reform around the governance of large social media platforms as well as oversight mechanisms, including regulations, on parent companies.

Though Facebook New Zealand, Google and Twitter all issued statements hours after the attack that they were working with the New Zealand Police to take down content associated with the attack, the content had by then spread far and wide across the web.

The video moved from platform to platform, edited, freeze-framed, downloaded off public streams which risked being taken down and then re-uploaded to private servers, which in turn served up the video to thousands more.

As Washington Post journalist Drew Harwell noted, "The New Zealand massacre was live-streamed on Facebook, announced on 8chan, reposted on YouTube, commentated about on Reddit, and mirrored around the world before the tech companies could even react".

4 https://www.stuff.co.nz/national/christchurch-shooting/111428947/terrorists-know-they-have-the-upper-hand-on-social-media

The challenge is significant because of the scale of the platforms, with billions of users each creating or consuming vast amounts of content every second. Managing the platforms is now largely algorithmic, meaning that only machines can cope with the scale and scope of content produced every second. There are serious limitations to this approach.

Terrorists know and now increasingly exploit it, weaponising the unending global popularity of social media to seed and spread their ideology in ways that no existing form of curtailment, containment or control can remotely compete with. And that's partly because of the way algorithms tasked with oversight of content are trained, which is entirely opaque.

It is entirely probable that algorithms trained to detect signs of radical Islamic terrorism are incapable of flagging a similar violent ideology or intent promoted in English, anchored to the language and symbolism of white supremacism or fascism.

In March 2018, Facebook's Chief Technology Officer (CTO) Mike Schroepfer noted that the company was using artificial intelligence (AI) to police its platform, and that it was "fairly effective" in distinguishing and removing "gore and graphic violence". Last Friday's killings highlight the risible falsity of this claim. Hours after the killings, dozens of videos featuring the same grisly violence as the original live stream were on Facebook.

One had generated 23,000 views an hour, with nearly 240,000 seeing it. Though Facebook notes it blocked 1.5 million videos in the days after the killings from being uploaded, it has tellingly withheld statistics on how many the original live stream reached or why 300,000 related videos were not identified soon after upload, which means they too were viewed – even for a short time – by hundreds of thousands.

And this isn't the first time graphic, wanton violence has resided on the platform for hours before it was taken down, by which time, the strategic aim and intention of producers has been met.

The problem doesn't end there. Neal Mohan, YouTube's Chief Product Officer, is on record saying how Christchurch brought the company's moderation and oversight to its knees.

It was unable to deal with the tens of thousands of videos spawned across its platform that showed the grisly killings – one every second at its peak. In two unprecedented moves for the company based on the severity of the challenge, his

team decided to block search functionality that allows users to search recent uploads and also completely bypass human moderation, trusting even with the possibility of false positives, content possibly linked to the violence in Christchurch flagged by its algorithmic agents.

Mohan has no final fix. The company just has no better way – even in the foreseeable future - to deal with another incident of this nature. Terrorists simply have the upper hand.

The Christchurch killer knew this and used it to his advantage. He won't be the last. The appeal to internet subcultures, famous personalities, memes, the very choice of music, expressions, gestures and popular references are a new argot of communications intentionally designed to use online cultures as means to amplify and promote violent ideology (called red-pilling).

At the same time, malevolent producers can almost entirely bypass existing controls and checks on the distribution of such material. The scale of social media is the hook, with the inability to oversee and inadequacies around governance, weaponised. Academics call this a wicked problem – a challenge that is so hard that even partial responses to any single aspect or facet increase the levels of complexity, often exponentially.

Generating greater friction around the production, promotion and spread of content is not in the interests of social media companies, who will continue to maintain – not without some merit - that billions of users producing vast amounts of mundane yet popular content daily is what primarily drives research and development. Read: profits.

Not without some irony, Facebook's Chief Operating Officer Sheryl Sandberg wrote in 2018 a glowing tribute to New Zealand's Prime Minister in Time magazine's list of 100 'Most Influential People'. After Prime Minister Jacinda Ardern noted that the live streaming of the grisly killings would be an issue she takes up with the company and perhaps mortified that this incident will strengthen calls around more robust regulation in the US, Sandberg had reached out after the violence, though it is unclear with what intent or assurances.

This rough sketch of the context I locate my doctoral studies in masks far greater complexity, anchored to community, culture, context and country. What is true of social media in Sri Lanka, my home and the central focus of my research, doesn't always hold sway in New Zealand. There are however strange parallels.

Repeated warnings around the weaponisation of Facebook to incite hate and violence, since 2014, went entirely unacknowledged by the company until severe communal riots almost exactly a year ago.

In Myanmar, the company's platforms were flagged by the United Nations as those that helped produce, normalise and spread violence against Muslims.

Till 2018, the company did little to nothing to address this, despite warnings and ample evidence from local organisations. YouTube's recommendation engine – the crucial algorithm that presents content that may interest you – has long since and openly been flagged as extremely problematic, beguilingly guiding users towards far-right radicalisation.

The Christchurch killer's release of a document before his rampage shows an acute understanding of how all this works, by transforming tired conspiracy into highly desirable material through strategic distribution just before an act that serves as the accelerant to infamy.

Alex Stamos, the former Chief Security Officer at Facebook, posted in the aftermath of Christchurch a sobering reminder of just why this violence goes viral. He notes that the language used, links provided and even excerpts of the violent video broadcast by news media only served to pique interest in the killer's original document and full video. This is a disturbing information ecology where content produced by terrorists cannot be taken down easily or quickly because the surge of interest generated around discovery and sharing will overwhelm attempts to delete and contain.

If this is the world we now inhabit and by using social media, contribute to cementing, the questions directed at companies and governments may be better directed at ourselves. How many of us searched for the video, and shared it? How many of us, without having any reason to, searched for, read and shared the killer's document? If we cannot control our baser instinct, then we become part of the problem.

The terrorists are counting on this, and us, to succeed. We should not let them win.

Pulse points

SCOOP NEW ZEALAND, 21 MARCH 2019[5]

Whether bound by country, city or community, the pulse of or, on Friday, the pain from a place like Christchurch can often be determined by the careful collection of social media updates published in the public domain. It is an interest in precisely this that brought me to New Zealand, where I study how Twitter and Facebook are integral to political communications and cycles of violence in Sri Lanka, my home. In South Asia, social media engagement drive attention towards or away from around key events, issues, individuals and institutions. Sport, religion, politics, elections and entertainment dominate content creation. The resulting conversations, to varying degrees, contest or cement opinions. Emotions drive engagement more than reasoned presentation or critical inquiry. Interestingly, though geographically distant and culturally distinct, a shared pattern of access and resulting behaviour on social media makes a younger demographic back home almost indistinguishable from their counterparts in New Zealand. This includes the heightened production of content on social media after an unexpected event.

Based on all this, I wasn't surprised to discover that the violence in Christchurch last Friday generated a tsunami of content just over Twitter. In the hours and days after the killings, specific hashtags on Twitter captured a community grappling with trying to make sense of, and recover from, a scale and scope of violence unprecedented in its history. The study of this content – much of it extremely painful to read - offers a glimpse into how the violence in Christchurch resonated access the country, and far beyond.

Almost immediately after the first news reports of the killings, #christchurchmosqueshooting, #christchurchshooting, #christchurchterroristattack, #newzealandterroristattack and #christchurch started to trend on Twitter domestically. This means that content using one or more of these hashtags showed a dramatic increase over a short period. In just a day, around 85,000 tweets featured one or more of these hashtags. By the 16th, two other hashtags started to trend - #49lives and #theyareus. In just a day, these two hashtags generated close to 37,000 tweets. With a single tweet capturing 280 characters, I was curious as to what just over 34 million characters, in the first 24 hours after the killings in Christchurch, said about the event. This is not just of academic interest. Policymakers and others interested

5 https://www.scoop.co.nz/stories/HL1903/S00124/pulse-points.htm

in or tasked with immediate response after a natural or man-made catastrophe can look at social media as a digital weathervane of public sentiment, crafting measures based on need, mood, reception or pushback.

When studied at scale, publicly shared content on social media is almost pathological. Key ideas, communities that assemble around specific individuals and content that goes viral can be gleaned through network science, which those like myself employ to understand key drivers and motivations behind content generation. This is easier to grasp by way of an example. Adil Shahzeb is in Islamabad, Pakistan and a television news presenter and host. And yet, on the 15th itself, he appears quite prominently in the content shared around the violence in Christchurch. This is, prima facie, utterly confusing. How can someone all the way in Pakistan become rapidly popular on Twitter around an event that happened in New Zealand? The answer is in a single tweet by Shahzeb, currently pinned to his Twitter profile, which identifies a man who tried to stop the killer as Naeem Rashid, with Pakistani origins. Rashid and his son Talha, the tweet noted, were tragically lost to the killer. This single tweet generated a considerable number of retweets and likes amongst those on Twitter, in both Pakistan and New Zealand. It is a similar story with Sunetra Choudhury, a Political Editor and journalist at NDTV, a popular Indian TV station. One of her tweets, featuring a clip of PM Ardern speaking to the affected community in Christchurch on Saturday, was viewed close to half a million times. The responses to the tweet, almost all from India, feature an overwhelming appreciation of the New Zealand PM's political leadership. These are two great examples of how empathy, shock and solidarity – here expressed in Urdu, Hindi and English – were able to cross vast geographies in a very short span of time.

Another way to get a sense of what's being discussed is to analyse the substance of the tweet. Through what's called a word cloud, words used more frequently can be rendered to appear larger than other words used less frequently. This process ends up with a visual map of the conversational terrain that affords the closer study of specific terms. Different hashtags feature different word clusters, but across all of them, Muslim, condemns, reject, Muslims, victims, terrorist, mentally, deranged, mosque, name, remembering, grotesque, white, supremacist and love feature prominently. The thrust, timbre and tone of tweets that feature these words are overwhelmingly empathetic and ranges from the profoundly sad to the outraged. By way of a loose comparison, when awful violence directed against the Muslim community broke out in Sri Lanka almost exactly a year ago, public sentiment I studied on Twitter at the

time didn't feature anything remotely akin to the levels of solidarity and support channelled towards the Muslim community in New Zealand, since last Friday.

What academics call a 'platform affordance' is more simply known to all Twitter users as a mention. Prefacing an account with the @ symbol ensures that on Twitter, a specific account is notified of a tweet. This is also used to direct a tweet towards a specific recipient or group. Unsurprisingly, PM Ardern, the Australian PM, the American President and controversial Australian Senator Fraser Anning are amongst those referenced the most over the first 24 hours. #49lives started trending on the 6th, generating nearly 17,000 tweets in a single day. The instigator of the hashtag is American. Khaled Beydoun is a Professor of Law based in Detroit, Michigan and a published author on Islamophobia. It is perhaps this academic interest that drove him to create #49lives, reflecting the number that at the time was the official toll of those killed in Christchurch. Beydoun's tweet, pinned to his profile, has generated an astonishing level of engagement – from New Zealand as well as globally. Liked nearly 146,000 times, retweeted just over 89,000 times and generated around 1,700 responses to date, the tweet prefigures PM Ardern's assertion in New Zealand's Parliament that she will not ever speak the killer's name. "I don't know the terrorist's name. Nor do I care to know it." avers Beydoun's tweet, which also asks to remember stories around and celebrate the lives of the victims. #theyareus generated just over 20,000 tweets by the 16th, but the sentiment or phrase is anchored to a tweet by PM Ardern made on the 15th. In a tweet liked 132,000 times and retweeted 40,000 times to date, she noted that "many of those affected will be members of our migrant communities – New Zealand is their home – they are us." However, it was two heartfelt tweets by Sam Neill, a businessman from Central Otago, that kick-started the hashtag trend. Speaking out against white supremacism and in solidarity with the Muslim community in New Zealand, Neill's two tweets, published consecutively on the 15th and 16th, have cumulatively generated nearly 27,000 likes, 4,200 retweets and 300 responses to date.

In sum, a cursory top-level study of the nearly 85,000 tweets generated in the 24 hours after the violence on Friday shows a global community outraged or dismayed at terrorism, an outpouring of love, empathy and solidarity, engagement that spans many continents and languages, addressing prominent politicians and journalists, featuring hundreds of smaller communities anchored to individuals based in New Zealand and beyond tweeting in a manner overwhelmingly supportive of the Muslim community.

The Twitter data underscores the value of studying public sentiment on social media in the aftermath of a tragedy. Social media provides pulse points. Framed by moments in time and driven by an understanding of, amongst other things, context, technology, access and language, the study of content in the public domain often helps in ascertaining how violence migrates from digital domains to physical, kinetic expression. Christchurch offers the world another lesson, a glimpse of which I wanted to capture here. Just as social media helps extremist ideology take seed and grow, it also helps in healing, empathy, gestures of solidarity, expressions of unity, the design of conciliatory measures and the articulation of grief and sympathy. The admiration, bordering on adulation, PM Ardern has received since Friday for her political leadership on just Twitter alone indicates that New Zealand is already seen as a template for how a country can and should respond to terrorism. These are more than just ephemeral in nature. Long after the world has moved on to the next news cycle, domestic conversations around what happened in Christchurch will endure on social media. Understanding how these ideas, anxieties and aspirations grow and spread lie at the heart of measures, over the long-term, that address extremism, racism, terrorism and prejudice, in all forms.

Principles over promises: Responding to the Christchurch terrorism

SCOOP NEW ZEALAND, 2 APRIL 2019[6]

Almost exactly a year ago, Facebook was in the news in New Zealand over a row with Privacy Commissioner John Edwards. The heated public exchange between Edwards and the company took place in the context of the Cambridge Analytica scandal, in which the private information of millions of Facebook users was harvested, illicitly, for deeply divisive, partisan and political ends. Edwards accused the company of breaching New Zealand's Privacy Act. The company responded that it hadn't and that the Privacy Commissioner had made an overbroad request which could not be serviced. Edwards proceeded to delete his account and warned others in New Zealand that continued use of Facebook could impact their right to privacy under domestic law. Just a few months prior, the COO of Facebook, Sheryl Sandberg, was pictured on Facebook's official country page with New Zealand PM Jacinda Ardern. The caption of the photo, which captured the two women in an embrace after a formal meeting, flagged efforts the company was making to keep children safe.

It is not surprising that Sandberg also wrote the paean to Ardern in last year's Time 100 list of the most influential people.

The violence on the 15th of March in Christchurch dramatically changed this relationship. In response to the act of terrorism, Facebook announced, and for the first time, a ban on "peace, support and representation of white nationalism and separatism on Facebook and Instagram". Two weeks after the killings in Christchurch, a message by Sandberg was featured on top of Instagram feeds in the country and featured in local media. The message noted that Facebook was "exploring restrictions on who can go Live depending on factors such as prior Community Standard violations" and that the company was "also investing in research to build better technology to quickly identify edited versions of violent videos and images and prevent people from re-sharing these versions." Additionally, the company was removing content from, and all praise or support of several hate groups in the country, as well as Australia. Sandberg's message called the terrorism in Christchurch "an act of pure evil", echoing verbatim David Coleman, Australia's immigration minister, in a statement he made

6 https://www.scoop.co.nz/stories/HL1904/S00012/principles-over-promises-responding-to-terrorism.htm

after denying entry to far-right commentator Milo Yiannopoulos, who after the attack referred to Muslims as “barbaric” and Islam as an “alien religious culture”. Last week, New Zealand’s Chief Censor David Shanks, declared the document released by the killer as ‘objectionable’, which now makes it an offence to share or even possess it. Following up, authorities also made the possession and distribution of the killer’s live stream video an offence. Facebook, Twitter and Microsoft have all been to New Zealand in the past fortnight, issuing statements, making promises and expressing solidarity. Silicon Valley-based technology companies are in the spotlight, but I wonder, why now? What’s changed?

Since its debut in 2015, a report by BuzzFeed News published in June 2017 flagged that at least 45 instances of grisly violence including shootings, rape, murders, child abuse and attempted suicides were broadcast on Facebook Live. That number would be higher now, not including Christchurch. The Founder and CEO of Facebook, Mark Zuckerberg, in May 2017, promised that 3,000 more moderators, in addition to 4,500 already working, would be added to over live and native video content. Promises to do more or better are what Zuckerberg and Sandberg are very good at making, in the aftermath of increasingly frequent and major privacy, ethics, violence or governance related scandals Facebook is in the middle of. Less apparent and forthcoming, over time, is what really the company does, invests in and builds.

There are also inconsistencies in the company’s responses to platform abuses. In 2017, the live video on Facebook of a man bound, gagged and repeatedly cut with a knife, lasting half an hour, was viewed by 16,000 users. By the time it was taken down, it had spread on YouTube. A company spokesperson at the time was quoted as saying that “in many instances… when people share this type of content, they are doing so to condemn violence or raise awareness about it. In that case, the video would be allowed.” Revealingly, the same claim wasn’t made with the Christchurch killer’s production.

The flipside to this is the use of Facebook’s tools to bear witness to human rights abuse. In 2016, the killing of a young black American Philando Castile by the Police in Minnesota was live-streamed on Facebook by his girlfriend, Diamond Reynolds, who was present with him in the car. The video went viral and helped document police brutality. There is also clear documented evidence of how violence captured from a Palestinian perspective, as well as content on potential war crimes, is at greater risk of removal on social media platforms. In fact, more than 70 civil rights groups wrote to Facebook in 2016, flagging this problem of unilateral removals based on

orders generated by repressive regimes, giving perpetrators greater impunity and murderers stronger immunity.

It is axiomatic that deleting videos, banning pages, blocking groups, algorithmic tagging and faster human moderation do not erase root causes of violent extremism. The use of WhatsApp in India to seed and spread violence is a cautionary tale in how the deletion of content on Facebook's public platforms may only drive it further underground. The answer is not to weaken or ban encryption. As New Zealand shows us, it is to investigate ways through which democratic values address, concretely and meaningfully, existential concerns of citizens and communities. This is hard work and beyond the lifespan of any one government. It also cannot be replaced by greater regulation of technology companies and social media. The two go hand in hand, and one is not a substitute for the other. It is here that governments, as well as technology companies, stumble, by responding to violent incidents in ways that don't fully consider how disparate social media platforms and ideologues corrosively influence and inform each other. Content produced in one region or country, can over time, inspire action and reflection in a very different country or community.

Take an Australian Senator's response, on Twitter, to the Christchurch terrorism. Though condemned by the Australian PM, the very act of referring to the Senator and what he noted on Twitter promoted the content to different audiences, both nationally and globally. The Twitter account, as well as the Facebook page of the Senator in question, produce and promote an essential ideology indistinguishable from the Christchurch killer's avowed motivations. It is the normalisation of extremism through the guise of outrage and selective condemnation. What should the response be?

In Myanmar, an independent human rights impact assessment on Facebook, conducted last year, resulted in the company updating policies to "remove misinformation that has the potential to contribute to imminent violence or physical harm". And yet, it is unclear how what may now be operational in Myanmar is also applied in other contexts, including in First World countries at risk of right-wing extremism.

I wonder, does it take grief and violence on the scale of Christchurch to jolt politicians and technology companies to take action around what was evident, for much longer? And in seeking to capitalise on the international media exposure and attention around an incident in a First World country, are decisions made in or because of New Zealand risking norms around content production, access and archival globally, on social media platforms that are now part of the socio-political, economic and

cultural DNA of entire regions? Precisely at a time when any opposition to or critical questioning of decisions taken on behalf of victims and those at risk of violence can generate hostility or pushback, we need to safeguard against good-faith measures that inadvertently risk the very fibre of liberal democracy politicians in New Zealand and technology companies seek to secure. An emphasis on nuance, context, culture and intent must endure.

So is meaningful investment, beyond vacuous promises. In 2016, Zuckerberg called live video "personal and emotional and raw and visceral". After the Christchurch video's visceral virality, it is unclear if Sandberg pushed this same line with PM Ardern. In fact, Facebook astonishingly allowed an Islamophobic ad featuring PM Ardern wearing a hijab, which was only taken down after a domestic website's intervention. Clearly, challenges persist. Social media companies can and must do more, including changing the very business models that have allowed major platforms to grow to a point where they are, essentially, ungovernable.

Grieving, we seek out easy answers. Banning weapons and blocking extremist content helps contain and address immediate concerns. Ideas though are incredibly resilient, and always find a way to new audiences. The longer-term will of the government to address hate groups, violent extremism in all forms and the normalisation of othering, from Maori to Muslim, requires sober reflection and more careful policymaking. What happens in New Zealand is already a template for the world. We must help PM Ardern and technology companies live up to this great responsibility.

A historic opportunity

OTAGO DAILY TIMES, 16 APRIL 2019[7]

New Zealand must get it right when it comes to legislating social media, writes Sanjana Hattotuwa

Wonderful news, said all the Sri Lankans. But why Queensland, all the Australians asked. Fifteen years ago, a Rotary World Peace Fellowship award offered seven universities around the world to undertake a Masters in Peace and Conflict Studies.

I chose the University of Bradford. I was awarded a place at the University of Queensland, in Brisbane. I didn't complain. The scholarship was a chance to get out of Sri Lanka and rigorously study what I had till then done on the ground, at a time when violent conflict dynamics were, after some years of relative calm, rising rapidly.

My Australian friends, however, were concerned that I would face in Queensland a degree of discrimination and intolerance they said I would never encounter in Sydney or Melbourne. I didn't know enough to argue and expected the worst.

After two years of extensive travel within the state and country, I returned to Sri Lanka experiencing very little along the lines I was warned about. Others though, at the same time, had a different experience - never physically violent, but far more verbally abusive. For them and I, this othering was at the margins of society. Well over a decade ago and without social media, violent extremism and ideology had to be actively sought after to be engaged with. Racism wasn't digitally dispersed.

It is with an enduring affection of Australia that I am deeply concerned about disturbing new legislation, passed hurriedly last week, which uses the terrorism in Christchurch to justify overbroad controls of social media.

The central focus of my doctoral research at Otago University is technology as both a driver of violence and a deterrent. How, today, social media promotes hate or harm is well known and widely reported.

As with any generalisation, though elements of truth exist, the simplification of a complex problem results in illegitimate targets of fear or anger. Social media companies, for their part, are irascibly unmoved by what for years those like me have

7 https://www.odt.co.nz/opinion/historic-opportunity

warned them about, around the abuse of platforms by those who seek to profit from violence.

Coherence and consistency in policies that respond to the seed and spread of violence are lacking and resisted. However, significant changes in stance, response and policies are coming.

The terrorism in Christchurch is responsible for accelerating globally what was sporadically mentioned or implemented with regards to safeguards around the production and promotion of content inciting violence, hate and discrimination. However, we must resist what appear to be simple answers to complex challenges, whether it comes from governments or big technology companies.

Violent extremism has many drivers, both visible and hidden. It doesn't bloom overnight. Social media, inextricably entwined in New Zealand's socio-political, economic and cultural fabric as it is back home in Sri Lanka, cannot be blamed, blocked or banned in the expectation that everything will be all right thereafter.

Driven by understandable concern around the dynamics of how the terrorism in Christchurch spread virally on social media, the Australian legislation - rushed through in just two days without any meaningful public debate, independent scrutiny or critical input - doesn't address root causes of terrorism, extremism or discrimination.

Among other concerns and though it sounds very good, holding social media companies and content providers criminally liable for content is a very disturbing template and precedent. American corporate entities are now required to oversee to a degree technically infeasible and humanly impossible, information produced on or spread through their services. This risks the imposition of draconian controls over what's produced, judged by hidden indicators, with little independent oversight and limited avenues for appeal. As a global precedent, the law is even more harmful, allowing comparatively illiberal governments to project or portray as the protection of citizens, parochial laws, essentially, that stifle democratic dissent.

David Kaye, the UN Special Rapporteur on the promotion and protection of the freedom of expression, is also deeply concerned. In an official letter to the Australian Minister of Foreign Affairs, Kaye stresses, among other more technical, procedural and legal points, the need for public review and proportionality, international legal obligations on the freedom of expression and imprecise wording in the law, which is entirely removed from how digital content is generated in society today, and by whom.

And herein lies the danger for New Zealand too. Politicians, under pressure to respond meaningfully, need to assuage the fears of a grieving country through demonstrable measures. The tendency is to pick an easy target and push through solutions that look and sound strong.

The underlying drivers of violence and conflict, however, simmer and fester. Measures taken to control and curtail gun ownership are welcome, and arguably, long overdue. Policymaking around social media, however, is a different problem set that cannot be as easily, or concretely, addressed.

This is not a submission to do nothing. Rather, it cautions against the understandable appeal of following the Australian response and law. Steps around the non-recurrence of domestic terrorism must certainly embrace aspects of social media regulation and related legislation.

The public must be involved in this. We know already that social media reflects and refracts - mirroring values of consumers as well as, through ways academics are struggling to grasp fully, changing attitudes and perceptions of users over time. This requires governments to iteratively work with social media companies on checks and balances that systemically decrease violence in all forms.

Elsewhere in the world, politicians who know the least about social media seek to control it, and those who know more or better, often abuse it.

Kiwis, led by PM Ardern's Government, have a historic opportunity to forge a response to terrorism - relevant and resonant globally - that incorporates how best government can work with technology companies to protect citizens from harm. Australia, with the best of intent, gets it very wrong.

New Zealand, with a greater calling, must get it right.

Can social media be a force for good in Sri Lanka after the Easter Sunday bombings? It's complicated

AUSTRALIAN BROADCASTING CORPORATION, RELIGION AND ETHICS, 25 APRIL 2019[8]

Writing about the inextricably intertwined relationship between social media and ethno-political violence in Sri Lanka is perversely easy — not least because there is such a mass of source material to draw upon.

Some of the first to understand the unprecedented power of audience capture and retention on social media were zealots, fanatics and racists. Mature misinformation campaigns that seeded and spread emotive content sowing anxiety, fear and division over social media existed as far back as 2014.

A comprehensive study — one of the first in South Asia — of content on Facebook inciting hate and violence against Muslims, despite being made widely available and in an easily accessible form, elicited neither response nor interest by the parent company. Two further reports, in 2015 and 2016, looked at content production around a high-profile court case involving an individual from the Sri Lankan Army and a key parliamentary election. In each study, hundreds of posts originally in Sinhala were translated into English. The translations could only be approximate because the venom of the vernacular often proved impossible to translate accurately into, or be captured adequately by, English.

Nonetheless, the three reports over as many years provide compelling snapshots of a social media platform in Sri Lanka that was already acting as a Petri dish for hate and violence. The most disturbing content exacerbated communal, ethnic, political and religious fault lines in the country. Arguably, these divisions had widened and deepened for decades prior to the entrenchment of social media. Before social media, traditional media married to, in the decades after independence in 1948, an increasingly toxic, dehumanising political culture provided the soil, over time, for communal tensions to develop into outright violence.

After the introduction of social media, malevolent actors weaponised the much celebrated democratisation of content production to construct, at scale and in a sustained manner, narrative frames that corrupted public conversation and

8 https://www.abc.net.au/religion/can-social-media-be-a-force-for-good-in-sri-lanka-after-the-eas/11045948

perceptions. The promise of social media as a domain that would hold traditional media and a rotting political culture accountable faded, as an information glut led to, counter-intuitively, a reduction in the spectrum of content to which those on social media were exposed, that they could consume and with which they could engage. Rather that alert users to diversity and difference, social media invisibly and purposefully averted their gaze to a grand illusion of choice that in fact was a spectrum entirely bounded by subjective bias and prejudice.

Content on Facebook in these early years bore an uncanny resemblance to what was being produced in Myanmar. In both countries, extremist Buddhist monks and allies were producing content viciously targeting Muslims in particular. The divisive narratives, the expressions, the symbols and graphics employed, the way targets of hate were being depicted through images and cartoons shared such shocking similarities that in workshops conducted in Myanmar, participants shown extremist content from Facebook in Sri Lanka, without understanding a word of the language, accurately grasped the hateful nature of the message, and vice versa.

It was not until March 2018, when Sri Lanka suffered the worst rioting and communal violence in many years, that Facebook took notice of the way its platform was part of the problem. In an attempt to control the spread of content that inflamed further violence, the government blocked leading social media platforms and services, including WhatsApp and Facebook. Confronted with international headlines, Facebook hurriedly visited Sri Lanka to meet with senior political leaders. Coincidentally, this was taking place at the same time that Facebook founder and CEO Mark Zuckerberg was facing unprecedented scrutiny in the United States over the company's role in Myanmar's genocidal campaign against the Rohingya Muslims. For the first time, the company officially responded to an open letter calling for greater oversight of content and remedial measures to control the flow of misinformation on its influential platforms.

Since that time, repeated interactions with and significant, substantive input from Sri Lankan civil society have strengthened internal oversight mechanisms, community guidelines and reporting channels. Though this is a marked improvement from a baseline of near zero investment or interest up to early 2018, the demeaning, divisive and de-humanising content produced by vast and varied constellations of producers continue to pose a significant challenge for the company to address in a timely and effective manner.

However, nothing in what I've researched to date on social media, over the last six

years, comes close to providing explanation for, much less causal linkage with, the terrorist attacks that killed over 360 people on Easter Sunday. The worst online expressions of hate I've seen have called for the extermination of children, women and men. As a result, there are some who propose that the normalisation of violent content, driven either by ideology or political gain, radicalises society in ways that cannot then be easily controlled — a digital Frankenstein-monster waging war against imagined enemies. And yet, nothing I observed on Facebook or Twitter before the attacks, including pages and accounts monitored with some of the most racist commentary, framed anything remotely close to what the country has now endured.

In the immediate and enduring aftermath of the attacks, the heightened production of misinformation has sought to weaponise the chaos within government and the resulting trust-deficit by propagating rumours that heighten anxiety and fear. In an unprecedented context where the theatre of terrorist operations extended from coast to coast, and reports of the discovery of arms caches, or of IEDs at the airport, or the arrest of associates and the controlled detonations of car-bombs extended for days after Easter Sunday, the misinformation being sown was clearly aimed at exacerbating the sense of crisis and communal agitation – making worse what is already a volatile situation. This is perhaps why, for the second time is just over a year, the government of Sri Lanka decided to block key social media platforms.

On the flip side, and much less reported or understood, is the role, reach and relevance of social media in public, private and political life. Facebook and allied services like WhatsApp, Messenger and Instagram power conversations that are always on and producing exabytes of data. Much of this content is unsurprisingly pegged to gossip, entertainment or pets. During the Rajapaksa regime from 2005 to early-2015, social media was increasingly used as the only viable and available vector for democratic dissent. Investigative journalism and activists embraced social media both to capture inconvenient truths and to publish eye-witness narratives. Critical civic media initiatives leveraged this foundation of a growing user base. Domestic realities and international developments were increasingly perceived through social media's overwhelming agenda-setting power. Fast forward to 2019, and social media content constantly emerges from every nook and cranny of society, in a country where special Facebook and WhatsApp subscriptions from leading mobile service providers allow for these services to be used without data restrictions on very cheap monthly rental packages.

For Western readers, I cede that this may all seem complex and confusing. The exponentially increasing potential for misinformation to grow at a volume and

velocity is truly mind-boggling. Complex ecologies and information economies at play today give preference and priority to the spread of information without check, care or consideration. Under authoritarian fiat or with the prospect of its imminent return, there is much to gain from a news and media foundation resistant to state capture, censorship and control.

On the flip or darker side, users who are in effect citizens in a country with poor media literacy, subject to unbridled choice, increasingly believe and then act upon content that is compelling and viral in nature, but often not or very loosely based on fact. Social media will continue to feature hate and violence, but to what degree this content informs or influences audiences is open to debate. The intersection of all this is where corporate course correction, legislation, regulation, technical innovation, political guidance and media literacy education converge.

Ultimately, though, I'd still contend that the core issue in Sri Lanka isn't really about social media. Evident before, during and after the Easter Sunday attacks was a tragic familiarity with what continues to thwart our fullest democratic potential. The distressing and disastrous lack of political will and leadership.

Sri Lankan social media expert: 'We must confront what we fear'

ASIA MEDIA CENTRE, 23 SEPTEMBER 2019[9]

At a time when social media is under scrutiny for its role in tragedies like the Christchurch mosque shootings, governments must avoid knee-jerk reactions, writes Sanjana Hattotuwa.

The markets — or "pola" as we call them in Sri Lanka — my father took me to when I was a child were chaotic. Vendors shouting over each other. Buyers animatedly haggling. The sights, sounds and smells of produce and livestock were overpowering to the senses, especially when at waist-height to everyone else.

Social media today reminds me of the markets of my childhood. Those who shout the loudest often get the most attention. It's easy to get lost in the madness. The more outrageous an idea or item held up for public display, the more crowds throng around it. Like different sections for vegetables, fruits, fish and meat in a "pola", social media platforms or accounts are branded by what their authors say, stand up for, are opposed to, or want to see more of.

In less than a decade, the same platforms that were attributed with enabling popular uprisings and revolutions overturning dictatorships are now flagged as tools of authoritarians or the megaphones of populists. Social media has become a reservoir of toxicity. For the most part, Silicon Valley companies are responsible for this mess. I should know. Documenting what was then a rising tide of violence against Muslims, I published evidence of inciting hate, racism and violence on Facebook as far back as 2013. The company said and did nothing.

Sri Lanka's social media shutdown

For years in Sri Lanka and elsewhere, just as much as it has contributed to violence, social media has become entwined with activism against authoritarianism. However, with every tragic act of violence or terrorism, media and governments are drawn to framing social media as something defined by emotion over evidence and fears over facts. This is dangerous. In the guise of public safety and security, governments in

9 https://www.asiamediacentre.org.nz/opinion/rising-to-the-challenges-of-social-media/

countries with a democratic deficit, like mine back home in Sri Lanka, are primarily interested in social media regulation only to censor and stifle democratic dissent.

For example, following the Easter Sunday terror attacks, the government blocked social media in an attempt to curb the flow of rumours and misinformation. It was the longest block in history, lasting close to two weeks. At the time, I recorded a rapid escalation in frustration that soon gave way to seething anger. Cut off from victims and loved ones, social media users around the country used VPNs to bypass the blocks and began to suspect the government was more interested in stifling criticism and public knowledge of intelligence failures that led to the attack. Ultimately, the block stymied the spread of content that sought to heal, and resulted in an angrier public more susceptible to rumours and the weaponisation of grief.

Lessons from Christchurch

How can evidence around the complexity of social media help fight against extremism, and celebrate its potential to be used for positive change? This question inspired my study of close to 100,000 tweets in the week after the Christchurch massacre in March. Domestic and international media at the time were aghast at how an awful livestream and hate-filled document had rapidly spread online. Yet, I found a different story on Twitter, anchored to content and commentary markedly different from what a terrorist wanted to inspire. I found a large number of tweets were in Urdu and Hindi, from Pakistan and India, respectively. Why? Because when prominent news anchors broke the story of Indian and Pakistani lives lost in Christchurch, tens of thousands of others from both countries, in their own language, joined in expressing dismay, distress, care and concern for Kiwis. There was a near-total absence of hate or violence and an abundance of love.

Rising above knee-jerk reactions

The theatre of terrorism thrives on stoking emotions that generate reactions. Politicians are driven to do something — anything. However, rarely does anything useful or enduring result. Rising above knee-jerk reactions, Prime Minister Ardern's 'Christchurch Call' recognises the "internet's ability to act as a force for good", "fostering inclusive societies". It also calls for the closer study of online content, over time. These are helpful frames to gather evidence required for progressive policymaking.

Violence and hate are easy stories for media, if only because there's a lot of it around and easily found. Yet, it also can blind us to content that binds and heals, somewhat hidden but often, just as abundant.

Just like global warming has no easy solution pegged to a single source, effort or location, governments, private corporations and civil society will need to work together to address the harms of social media. This is why I argue for a more dispassionate study of data, framed by values entrenched in New Zealand's constitution, country, culture and communities. This is hard work, with no easy wins. But robust evidence and data on social media should inform official responses, reforms and regulations. I believe that confronting what we fear can help us identify the worst of what social media is, which is the first step towards making it a more civil environment. We must rise to the challenge.

Is Social Media Driving Instability?

NEW ZEALAND CLASSIFICATION OFFICE, 12 OCTOBER 2020[10]

What could I do to destabilise New Zealand? Quite a bit, as it turns out, given a few years and the unprecedented volatility brought about by the Coronavirus pandemic.

This question is far removed from personal desire or an easy endeavour with a guaranteed outcome. I ran it by the Classification Office during a visit as a thought-experiment based on on-going research looking at social media's role in helping spread anxiety, violence, and hate. What could be key drivers of instability? What could content aimed at stoking fear or anxiety look like? How could violence be encouraged, first digitally and then, justified physically? Could anyone find out what made Kiwis anxious, or kindled their interest the most, daily? What are the implications of this data when studied in the aggregate?

My work at the National Centre for Peace and Conflict Studies at the University of Otago seeks to answer some of these complex questions. I arrived in New Zealand after nearly two decades of advocacy and activism involving some form of internet or web-based technology in Sri Lanka. Before around 2012, the likes of Facebook and Twitter helped protect and promote content at risk of censorship or violent suppression. Over the past decade, however, social media platforms increasingly helped to seed and spread toxicity, hate and violence. What went wrong? There are many answers, but all revolve around the fact that leading social media companies in Silicon Valley simplistically assumed connecting people was a net good and democratic gain. They were wrong.

Research as far back as 2013 provided some of the first evidence globally of how Facebook was used by violent extremists to fuel Islamophobia in Sri Lanka. But it wasn't until March 2018, after the country's worst anti-Muslim riots in decades that Facebook was compelled to meaningfully investigate the role its products and platforms played in spreading the violence. Long overdue measures around non-recurrence were undertaken, but it was too little, too late. A toxic genie was out of the bottle.

Though thousands of kilometres away, tragic lessons from Sri Lanka matter to New Zealand. It turns out that while social media is engineered to encourage the sharing

10 https://www.classificationoffice.govt.nz/blog/is-social-media-driving-instability/

of what we feel, fear, desire or do, what we end up posting – over time and also in real time – is often used to seed doubt, sow anxiety and spread anger. How and over which platforms this is done changes, as does the effectiveness of various types of content. However, after the pandemic – globally as well as in New Zealand - there is significant anxiety about job security, unemployment, the economy, health, travel and the future. Sophisticated domestic and international actors and political entrepreneurs seek to amplify these concerns for selfish benefit or partisan gain. While clear to many policymakers and academics, the public struggle sometimes to understand the magnitude of the risk.

Imagine a day-care centre for children with just two staff named, purely for illustrative purposes, Facebook and YouTube. With a few children, all from the same neighbourhood and in a large room, Facebook and YouTube manage their roles just fine. They tend to the needs of the children, look out for deviant behaviour, help those in distress, watch out for risks and maintain healthy interactions amongst kids not too different from each other. There's little to no violence, and if there is an outburst, it is quickly addressed. Now imagine this day-care centre, over a decade, growing to fill a skyscraper. Children of varying ages and backgrounds fill each floor. However, it's still Facebook and YouTube managing all of them. Without any oversight, the children run wild. Anything goes, and without correction or guidance, bad behaviour has no consequence or even an established template for the achievement of certain ends. Even with the kids they can see, Facebook and YouTube are completely overwhelmed by competing needs. Without adequate resources and support, things quickly disintegrate into total chaos.

No day-care centre with such nightmarish under-staffing would be allowed to operate ethically or legally. And yet, this is not unlike the management and operations of leading social media companies, with billions of users. As with many other countries, New Zealand's national conversation is increasingly mediated through social media platforms or products, governed by companies struggling to deal with toxicity, hate and violence. This complex ecology is ripe for abuse and weaponisation, in ways tried and tested elsewhere. The danger, to my mind, lies in an exceptionalism that sees New Zealand as mostly immune to democratic decay through the slow but steady drip of toxicity over web, internet and social media.

I believe this country offers many lessons in progressive policymaking and the regulation of social media. Undoubtedly, it will be a challenging and lengthy process. With the pandemic response, we know how careful study, contact tracing and evidence-based analysis help locate, in a timely manner, sources of infection and

super-spreaders, helping the strategic containment of a deadly virus. The same principles can be applied to social media content around hate and harm. New Zealand is well-positioned to lead this process. Instruments like the Christchurch Call highlight how social media platforms can no longer wish away the harm they feature and, often, amplify. Sober responses to emotional, divisive issues in a media landscape that is continually evolving are hard to imagine, but essential to craft. My research contributes to what many others, in New Zealand and elsewhere, are doing to strengthen our better angels.

Ma tini ma mano ka rapa te whai.

Sanjana Hattotuwa is a PhD candidate at the University of Otago and Special Advisor at the ICT4Peace Foundation. His views do not necessarily represent those of the Chief Censor or of the Classification Office.

The Christchurch massacre and social media: Lessons learnt and unlearnt [11]

As part of the New Ec(h)o systems: Democracy in the age of social media, Paul Ash, the New Zealand Prime Minister's Special Representative on Cyber and Digital and Cyber Coordinator at Department of the Prime Minister and Cabinet joined Special Advisor Sanjana Hattotuwa in a session titled 'The Christchurch massacre and social media: Lessons learnt and unlearnt'.

Ash also heads the Christchurch Call. In addition to being on the Call's Advisory Network, the Foundation's Chair, Daniel Stauffacher was invited to civil society meeting hosted by the New Zealand Prime Minister Rt Hon Jacinda Ardern in Paris on 14 May 2019. The session also features a topline presentation of Hattotuwa's pioneering doctoral research in New Zealand, looking at content dynamics on Twitter after the Christchurch massacre, supported by the first of its kind partnership with Twitter's Data for Good initiative and the National Centre for Peace and Conflict Studies (NCPACS) at the University of Otago.

The session's framing

The events of 15 March 2019 changed New Zealand's approach to and study of violent extremism. Just over a month after, suicide attacks across Sri Lanka claimed over 5 times as many victims. Leading up to, during and in the aftermath of both incidents, social media played a significant role. The similarities end there. New Zealand's cri de coeur, the Christchurch Call, aims to reduce platform harms, including the spread of hate and violence. Social media was instrumentalised in Sri Lanka after the attacks to stoke Islamophobia. In both countries, however, episodic, preconceived media coverage glosses over more interesting developments. Where is the Call today? What is the platform's future? And in Urdu, Hindi, Turkish and Hausa, why did victims in Christchurch galvanise empathy in ways Sri Lankan victims did not? What lessons for platform governance can both countries offer?

Readers are invited to view the session on YouTube at https://ict4peace.org/activities/the-christchurch-massacre-and-social-media-lessons-learnt-and-unlearnt/

11 https://ict4peace.org/activities/the-christchurch-massacre-and-social-media-lessons-learnt-and-unlearnt/

Addressing the Infodemic

ICT4PEACE FOUNDATION, 6 JULY 2021[12]

Meaningful policymaking to fight the swelling seed, spread and supremacy of misinformation benefits from data and evidence alive to socio-political realities. A new report from Aotearoa New Zealand's Classification Office does precisely this. 'The Edge of the Infodemic' presents a snapshot of the country's media and information ecologies in addition to the risk of misinformation, which I fear risks expansion and entrenchment.

Headlining the findings, 82% of New Zealanders are concerned about the spread of misinformation. In comparison, 75% think false information around Covid-19 is a significant threat to society. The report also finds those who consume and trust online-only sources of information are more likely to believe false statements. For researchers studying misinformation, the report by the Classification Office holds no surprises.

Several tensions emerge in the report, ranging from the consumption of news and information to what New Zealanders want done to stem or stop the flow of misinformation. At the outset, codified regulation and legislative instruments are inadequate to deal with the problem even as it stands today; leave aside how misinformation will metastasise in the future. Aotearoa New Zealand's geographic isolation will not inoculate her against misinformation's progressively corrosive effects on democratic institutions, electoral integrity, social cohesion, political culture, and media standards. However, meaningfully responding requires familiarity with a vocabulary of harmful or violent intent that many, including in government, may not have. A rough analogy would be to ask firefighters who have only trained in urban scenarios to douse a forest fire. They know what they see but know not how to control or contain the spread of flames in a context fundamentally different to what they were trained for. Policymaking in Aotearoa around misinformation faces a similar challenge.

The report brings out an interesting tension between a general distrust of online sources of information and, simultaneously, social media and the internet emerging as the most popular sources of information. Facebook apps or platforms, including Messenger and Instagram, are predominant news and information vectors, aside from

12 https://ict4peace.org/activities/publications/foundation/addressing-the-infodemic-2/

YouTube. As a high-trust society, even with the increasing seed of misinformation in Aotearoa, the subsequent spread is contained by society's ingrained scepticism of scurrilous or spurious claims. However, because of what researchers see in other contexts, heading to Aotearoa New Zealand with a vaulting ambition, this resilience cannot be taken for granted for at least two key reasons.

First, misinformation's harms and study do not often consider the selective capture of issues by political entrepreneurs and the vocabulary used to promote them. The justification of racism, persistent denial of Te Tiriti, the othering of Maori and ostracising minorities, all present in mainstream political discourse, is misinformation that isn't recognised as such when the study of it is limited to specific categories linked to Covid-19 or other conspiracy theories. In different contexts, violent othering and dehumanising language normalise divisive definitions and polarising opinions. This drip-feed of divisive frames can, over a shorter period than many expect, fundamentally change the nature of society.

Secondly, the complicated role of social media platforms simultaneously (and often confusingly) contributes to misinformation's spread and measures to contain it. As much as social media platforms have served to globalise movements like Black Lives Matter or School Strike for Climate, they are also persuasive, transnational highways for hate and harm. Misinformation that's freely and easily accessible on social media supports a compelling worldview for those who feel underrepresented, misrepresented, absent or violently excluded. Perhaps surprising strong advocates of media and information literacy, the presentation of facts or corrective measures to address conspiratorial perspectives, especially by experts or those associated with government, counter-intuitively risk cementing the very beliefs they seek to dislodge.

Today, conspiracies like QAnon, first grown in and for the US, quickly and easily travel to and find domestic resonance in Germany, the UK, France, and Aotearoa. Aiding this toxic transfer is the absent or uneven application of guidelines against misinformation by social media companies. 84% of New Zealanders want a range of actors, including government agencies, officials, experts, news media and social media companies, to address misinformation. What's unclear in the Classification Office report is just what this will entail, given that options range from voluntary codes of conduct to government regulation, revised laws, more robust platform governance, media and information literacy and the deplatforming of misinformation super-spreaders. Existing regulatory frameworks and policy frameworks are no longer fit for purpose to address these complex challenges.

Notably and as a norm-setting example, 'The Edge of the Infodemic' responds to a key recommendation by the Royal Commission on the Christchurch massacre, calling public sector agencies to collect data that helps develop policies strengthening social cohesion. As a snapshot of a post-pandemic country, it is unclear if and to what degree Covid-19 contributed to the significant concern around misinformation. However, given a problem that isn't going away, survey research of this nature should be conducted annually as a barometer of socio-political sentiments, anxieties, perceptions, and behaviours. For example, pertinent to current debates on the revision of hate speech legislation, 'The Edge of the Infodemic' highlights that 79% of New Zealanders disagree that people and organisations should be able to say what they want on social media even if it leads to violence or self-harm. Just 22% think there should be no limits on what people say.

Conversely, Aotearoa New Zealand is almost evenly split between those who believe people and organisations should be able to say what they want even if it offends or upsets others, and those who disagree. On the face of it, this suggests that moves to criminalise misinformation will not work or find broad acceptance, as the Chief Censor's introduction to the report clearly notes. Instead, guardrails against democratic decay require measures to strengthen trust in Aotearoa's news media, government agencies, officials, and her experts, clearly brought out in the report.

In contrast to more violent contexts I study, Aotearoa New Zealand has a window of opportunity to meaningfully address challenges flagged in 'The Edge of the Infodemic'. However, it may not last long. The Royal Commission Report's stress on social cohesion is central to policymaking in this regard. In how and to what degree Aotearoa gives life to the legislative and regulatory expression of aroha (empathy, compassion), manaakitanga (respect and care for others) and mōhiotanga (insight and comprehension) in fighting misinformation will lie the success of the endeavour.

Acknowledged in the Classification Office report, Sanjana provided input into the questionnaire design as well as early drafts.

The nature and nurture of disinformation

ICT4PEACE FOUNDATION, 7 JULY 2021[13]

For us the land is matrix and destroyer,
Resentful, darkly known
By sunset omens, low words heard in branches.
— Poem in the Matukituki Valley, James K. Baxter

"In the woods we return to reason and faith."
— Ralph Waldo Emerson

Aotearoa New Zealand's inaugural hui on countering terrorism and violent extremism, He Whenua Taurikura, meaning a country at peace, was held mid-June in Christchurch. A policy brief written for and a presentation made at the hui articulated a radical departure from mainstream thinking on how best to address disinformation. Knowingly spreading false or misleading information is a significant challenge that will only grow in complexity, including in Aotearoa. Populists and their propaganda have utilised a corrosive choral of partial, partisan content for decades. More recently, social media is an accelerant to infamy, potentially expanding the reach of harmful content not just across county or country, but continents. Extremely complicated and, by its very nature, constantly changing, disinformation's entrenchment in Aotearoa New Zealand demands policies that are fit for purpose, responsive, sensitive to local context and grounded in her culture.

Departing from disinformation defined as predominantly an online or social media phenomenon sought to focus on how to address it meaningfully, inspired by two key factors. Firstly, research, spanning two decades, studying the seed, spread, and settling of harmful content in divided societies. Secondly, and new to existing approaches, Aotearoa New Zealand's salubrious natural environment as the canvass for an ecological perspective to address the root causes of disinformation.

Debate around proposed revisions to laws governing hate speech in Aotearoa New Zealand reaffirm the need to step back and reflect on whether proposed solutions adequately embrace the complexity of problems facing democracy, and how these interconnected challenges will change. Supported by research looking at the criminalisation of misinformation elsewhere in the world, there is scant evidence to

13 https://ict4peace.org/activities/publications/the-nature-and-nurture-of-disinformation/

support the belief that punitive measures help meaningfully stem the flow of harmful content, or disincentivise the production of disinformation. Furthermore, mirroring what virology calls the gain of function, existing pathways for disinformation are constantly mutating, deflecting oversight and meaningful, timely responses by social media companies, regulators, governments or academics. Accordingly, predominantly bureaucratic or technocratic approaches risk establishing regulatory or legislative frameworks that inadequately address issues they seek to address. Governments are often motivated by the need to do something, and be seen to be doing something, around pressing socio-political problems constituents demand responses to. However, disinformation operates on a different timescale. While headline-grabbing moments like the horrific violence on 6 January in Capitol Hill focus attention on online disinformation's increasing contribution to offline violence, harm can be fomented and fester over a much more extended period through toxic content that, for years, escapes scrutiny. This is mainly because disinformation's highly motivated, chief producers today are far removed from ignorant, basement-inhabiting stereotypes and demonstrate significant skill, strategy, and sophistication in the production of harmful content. Their principal motivation is to radically change societal attitudes, practices, and perceptions over time, often to become more exclusive, partial, monotone, and insular. Accordingly, while disinformation may be increasingly digital in form, its function is often to foment and fan unrest, offline.

At He Whenua Taurikura, I used the analogy of how the Department of Conservation would go about caring for and protecting Aotearoa New Zealand's natural environment as a radically new frame to address disinformation. Instead of seeing regulation and codified laws as the most effective way to stop disinformation's seed and spread, I flagged how the management of forests and parks focussed on ecological balance and inter-dependencies. In nature's rich, diverse ecologies, the quality of groundwater, soil and underbrush directly influence the health of a forest or park. What grows or does not is in a symbiotic relationship with the larger environment. Pruning branches, cutting trees or new saplings will not lead to a healthier ecology if groundwater is contaminated, or what's planted is incompatible with native wildlife. The same applies to laws, regulations and policies around content inciting hate, where what's proposed and implemented needs to be fit for purpose, country, context, community and culture. A whole of society approach to combat disinformation – what the Royal Commission Report on Christchurch terms 'social cohesion' – requires us to address, amongst other issues, the enduring violence of post-colonial history, unmet Te Tiriti obligations, and acknowledge other domestic problems, like racism. While disinformation may never be eliminated, social and political responses to content

inciting hate and harm can and should be cultivated in such a way as to reject calls for or expressions of violence.

To this end, a slide featuring the movement of a flock of birds, called murmuration, captured the audience's imagination at the hui. The study of disinformation amongst large groups, I noted, was very similar to the way shoals of fish or flocks of birds responded to their environment, including by signalling the presence of predators only or first detected by a few at the edges. Groups tend to gravitate towards, cluster around or reject things in ways that, studied at scale, can help create more robust defences against predatory, violent or hateful actors. Strengthening belonging, inclusion, participation, recognition and legitimacy as a whole of society approach to minimising disinformation's harms build on the Royal Commission Report's recommendations. The UN Special Rapporteur on the Freedom of Expression Irene Khan makes a similar point in a recent report, noting that,

> *"Disinformation is not the cause but the consequence of societal crises and the breakdown of public trust in institutions. Strategies to address disinformation are unlikely to succeed without more attention being paid to these underlying factors."*

Going beyond purely technocratic or predominantly algorithmic approaches, I propose drawing from conservation, behavioural sciences, biology and ecological studies to address disinformation's nature and nurture. A more holistic and grounded approach also draw from from Aotearoa New Zealand's unique bi-cultural values, that Foreign Minister Nanaia Mahuta recently highlighted in her speech at the 55th Otago Foreign Policy School, including manākitanga (a common humanity) and whanaungatanga (connectedness).

The lines from Baxter quoted at the start highlight nature's destructive potential, like disinformation's harmful patina on society, around which there are already abundant warnings. On the other hand, Emerson enjoins us to return to nature in order to learn how best to address challenges like disinformation, for which there is, and never will be, simple solutions.

PODCASTS

Social media: promoter of democratic participation or purveyor of violence?

AUSTRALIAN BROADCASTING CORPORATION, RELIGION AND ETHICS, 24 APRIL 2019[14]

In the wake of a deadly series of coordinated bombings that tore through churches and hotels across Sri Lanka on Easter Sunday morning, claiming more than 300 lives and injuring 500 others, the Sri Lankan government blocked access to Facebook and its related services, WhatsApp and Instagram, as well as YouTube. For the second time in little more than a year, the government has deemed it necessary to ban these social media platforms in order to prevent the uncontrolled spread of misinformation, rumours, conspiracies and outright lies. There is no denying the role that this unrestrained, instantaneous dissemination of 'news' has played in the eruption of ethnic and religious violence, not just in Sri Lanka, but in Myanmar, Indonesia, India and Mexico.

And now that Islamic State has claimed responsibility for these attacks (legitimately or not) and have tried to connect them to the Facebook-mediated mass shootings in two mosques in Christchurch, New Zealand (again, rather dubiously) — thereby ensuring that the Easter Sunday massacre will be enlisted in one or another narrative of violence and victimhood — the likelihood of reprisals has increased exponentially, particularly against the already vulnerable Muslim minority in Sri Lanka.

The unavoidable question is: Can social media platforms, like Facebook, be anything other than purveyors of violence in countries with deep and long-standing histories of ethnic-religious tensions? Has Facebook's deliberately rapid and overtly opportunistic expansion into future 'markets' foreclosed the possibility of concerted reflection on whether its service is a social good in those countries? Are there attendant risks, in countries without robust public institutions and high levels of civic trust, in blocking access to unmediated, uncensored forms of digital communication?

Duration: 44min 48sec
Broadcast: Wed 24 Apr 2019, 11:30am

14 https://www.abc.net.au/radionational/programs/theminefield/social-media:-promoter-of-democratic-participation-or-purveyor/11042176

PUBLIC LECTURES

Full video & slidedeck of lecture: From Christchurch to Sri Lanka – The curious case of social media

ICT4PEACE FOUNDATION, 17 JUNE 2019[15]

On 20 May 2019, Sanjana Hattotuwa, a Special Advisor at the ICT4Peace Foundation since 2006, gave a well-attended public lecture at the University of Zurich on the role, reach and relevance of social media in responding to kinetic and digital violence, including the potential as well as existing challenges around artificial intelligence, machine learning and algorithmic curation. The lecture was anchored to on-going doctoral research, data-collection and writing on the terrorist attacks in Christchurch, New Zealand in March and the Easter Sunday suicide bombings in Sri Lanka – Sanjana's home.

Sanjana's presentation started with an overview of his doctoral research and scope of data-collection, anchored to Facebook and Twitter in particular. The daily capture and study of this data gives him perspectives at both a macro-level (quantitative) and with more precise granular detail (qualitative) which help in unpacking drivers of violence, key voices, leitmotifs and other key strains of conversations on social media after a violent incident. Comparing the terrorist incidents in Christchurch and Sri Lanka, Sanjana contextualised the global media coverage around both incidents and in particular, the criticism against social media following the live-streaming of the Christchurch incident. Calling social media an 'accelerant for infamy', Sanjana proposed an original thesis around how the science of murmuration and the study of mob mentality (based on the three key principles of adhesion, cohesion and repulsion), when applied to conversational and content related dynamics online, could provide insights into how violence spread and generated new audiences.

Sanjana then spoke about artificial intelligence (AI), and despite the more common framing by mainstream media, significant challenges around AI-based content curation faced by leading social media companies at present. Aside from Facebook and Twitter, Sanjana flagged the extremely problematic recommendation engines of

15 https://ict4peace.org/activities/full-video-slidedeck-of-lecture-from-christchurch-to-sri-lanka-the-curious-case-of-social-media/

YouTube, including the recent misrepresentation of the Notre Dame fire. Using two images, he also flagged how even the simplest of manipulation still baffled the most sophisticated of AI, looking at image classification (which is central to the identification of violent or hateful content online).

Using Sinhala – a language spoken only in Sri Lanka – Sanjana highlighted the challenges of natural language processing (NLP), which akin to AI, was central to content curation at scale. In one slide, he typed Sinhalese characters and in another, showed an image with characters embedded into it, noting something different to what was typed. Sanjana noted that the first, by itself, presented a number of challenges for companies that had for too long ignored the likes of Sinhalese or Burmese content generated on their platforms, while the second compounded those issues, by presenting to AI and ML architectures nuance, context and script training datasets at present aren't based around or on.

Sanjana then went on to explain the dangerous consequences of 'context conflation' in a country or context with very high adult literacy and very poor media literacy. While some or all of this is known, Sanjana went on to then frame and focus, through hard data, the manner in which Twitter provided a global platform after the violence in Christchurch for people to discuss solidarity, express sadness and generate strength. A conversation far removed from hateful right-wing ideology or the promotion of violence took place, rapidly and vibrantly, on the same social media platforms that the global mainstream media chastised for having played a central role in the promotion of violence. Noting the unprecedented constitutional crisis in Sri Lanka late-2018 as well as the content produced after the Easter Sunday attacks, Sanjana again highlighted how social media in general, and Facebook and Twitter in particular – played a central role in democracy promotion, dissent, activism, pushback against authoritarian creep and the promotion of non-violent frames after a heinous terrorist attack.

Sanjana ended the lecture by looking at inflexion points – noting that social media companies, civil society and governments needed to recognise a historic opportunity to change the status quo, including core profit models and business practices, in order to ensure to extent possible social media didn't provide ready platforms for fermenting or fomenting fear, hate, violence and terrorism.

Sanjana underscored why the #deletefacebook movement in the West would never take root in countries like Myanmar or Sri Lanka, and was a risible suggestion to hundreds of millions using Facebook's spectrum of apps and services. He noted

the dangers around the emulation, adaptation or adoption of regulation from the West in countries with a democratic deficit, while at the same time noting the importance of regulation's introduction to govern companies that needed oversight to a greater degree than is present today. Linked to this, he noted that Silicon Valley's business models were anchored to quantity over quality, and the generation of content irrespective of the timbre or tenor of that material – leading to the obvious weaponisation of platforms never meant to be Petrie dishes for terrorism and violent ideologies. Facebook's recent pivot to privacy, announced by Mark Zuckerberg, Sanjana welcomed with cautious optimism, noting that while there was much to celebrate and welcome, it could also mean that academics would find it much harder or downright impossible, in the future, to study the generation and spread of violent extremism on social media. Sanjana spoke about social media as being central to the DNA of politics, political communication and social interactions in countries like Sri Lanka, noting that as a consequence, there is no alternative to the development of AI, ML and NLP techniques to deal with the tsunami of content generation growing apace, every day, already far beyond the ability of a few hundred humans to oversee and respond to. In both the penultimate and final slides, Sanjana spoke to the need to problematise the discussion of media and social media, noting how complex a landscape it really was, defying easy capture or explanation.

The lively and interesting Q&A session, which exceeded the allotted time, went into a number of aspects Sanjana touched on. The video above captures the Q&A segment as well.

POLICY BRIEFS

Re-imagining responses to extremism: The importance of context, culture and community [16]

ICT4Peace Foundation's Special Advisor Sanjana Hattotuwa was invited by New Zealand's Department of Prime Minister and Cabinet to write this policy brief on the occasion of He Whenua Taurikura, New Zealand's first annual hui (meaning a large gathering in Maori) on countering terrorism and violent extremism. The hui was held from 14-16 June 2021 in Christchurch. He Whenua Taurikura translates to 'a country at peace'.

###

Aotearoa, New Zealand will face increasingly sophisticated campaigns to seed and spread anxiety, fear and anger, both online and offline. These campaigns will emerge from or be amplified by political entrepreneurs from within the country and outside it. Inoculation against this democratic erosion – such that it exists at present – risk diminishing returns over time in the face of iterative, intentional and

16 https://ict4peace.org/activities/policy-brief-on-re-imagining-responses-to-extremism-the-importance-of-context-culture-and-community/

unrelating "everyday campaigns" across a range of issues, including but not limited to partisan politics, proposed and existing laws, bi-cultural relations, health, elections, infrastructure and jobs. As the scope, scale and speed of disbelief grows, trust in democratic institutions, including electoral outcomes, will decrease. No electorate is immune, and what is a possible future scenario for Aotearoa, New Zealand is well entrenched in other countries which are now templates for engineering democratic deficit.

The long-game of anti-democratic architects is to weaponise scepticism. Like a digital Novichok, the manner in which society sees itself, negotiates difference, communicates with each other, deals with the past, and envisions the future – and an individual's or community's place in it or ownership of it – can be corrupted through online content and social media platforms. Unlike a nerve agent however, which has an immediate and visible physiological impact, through influence operations conducted over time, the tone, timbre and thrust of divisive frames can become the foundations of political and social discourse. Sociologist Diane Vaughan called it "the normalisation of deviance" in relation to what caused the Challenger Space Shuttle disaster in 1986. Over time, individuals can come to accept a problem as a feature, instead of an aberration. The bad actors become those amongst us – our extended family, friends and neighbours – who come to believe in things we can no longer identify with, or subscribe to. It is, ultimately, the weaponisation of PM Jacinda Ardern's "They Are Us", through the strategic, systematic and sustained dismantling of democratic ideals, institutions and processes. Without a consistent, clear or common enemy, existing strategies to safeguard Aotearoa, New Zealand from democratic decay risk failure, and at a pace quicker than many in government, media and civil society expect or plan for.

The perspectives in this policy brief are informed by two inter-related drivers – one, the lived experience of negotiating violent conflict in Sri Lanka since 2002, including responding to online manifestations of offline violence for over a decade and, two, doctoral research looking specially at the role, reach and relevance of Facebook, Twitter and social media in simultaneously fuelling and quelling socio-political violence. This research included how online content is inextricably entwined with and informed by offline developments including but not limited to communal riots, significant political unrest, high-casualty terrorism, and consequential electoral moments.

The point I seek to stress is a simple one. Coming from, and calling home a country that is, in every imaginable way and every day, profoundly more violent than Aotearoa, New Zealand in most touch points for citizens, and especially those from minority

communities, I viscerally appreciate the symbolic invocations and implications of statements by political entrepreneurs or their proxies. Sometimes called dog-whistling, the reach and resonance of references intended for specific audiences is a code that if and when cracked, provides vital insights into intent, motivation and strategy of despotic innovation. However, echoing what Polish-American scientist and philosopher Alfred Korzybski's remark that "the map is not the territory", disinformation's social and political impact is more complicated than just the study or presentation of big data.

Data can help show us what's going on, but not unlike Rorschach blots, resulting visualisations only make sense when read in specific contexts. Words like online extremism and digital world tend to project violence as predominantly determined by digital content. The telos of this gaze – which has served democracies well but is no longer fit for purpose – is to see legislative instruments, laws, the codification of regulations and punitive measures as adequate, desirable or definitive responses for disinformation's Hydra-headed entrenchment, expanding at pace. Informed by lived experience, activism, and research, I study online data in situ, seeing digital interactions as inextricably entwined with local cultures, histories, communities, media ecologies, political cultures, anxieties and aspirations.

Consequently, I argue that disinformation goes to the heart of who we are, what we believe in, love to do, and why. It is an existential inquiry and exercise, not (just) a digital study or phenomenon. By its very nature, disinformation is socio-technological, being offline in nature as much as it is increasingly online in nurture. It follows that disinformation requires systems or lateral thinking to grasp, beyond technocratic or bureaucratic frames. While appreciating their role, I argue that we must be sceptical of all legal or legislative responses to what are essentially, and will remain, socio-political problems present in online and offline forms, simultaneously.

Why is an inter-disciplinary, broad spectrum approach vital to safeguard democracy? Even as legislators seem convinced they have a handle on fake news and hate speech definitions, researchers grapple with the morphology of content inciting hate and violence. Hate, harm and violence are, in fact, often very hard to assess. Digital content is iterative and requires contextual knowledge to understand the implications of. Cross-pollination is the norm, where engagement on one platform leads to variations of the content and commentary on another – an inosculation that sees digital hate grow in tandem to offline developments. With each opportunistic migration from one app, platform or vector to another, frame, function and form of content changes.

The speed, scale and scope of this migration and morphing has long overtaken the imagination of policy makers, most regulators and even social media companies, resulting in an everyday tsunami of content that defies meaningful oversight or rapid response. Furthermore, ambiguity is now a strategic choice, where content that resides right at the borderline of what's prohibited by social media platform serve as sufficient signals for followers to amplify specific messages, including targeted hate. Political and media entrepreneurs in the Global South are now joined by those in Europe, US and Global North in instrumentalising social media platforms as bully-pulpits or manic megaphones.

This pulsating pathology of disinformation – that's far more complex than this snapshot – already resides within and outside Aotearoa, New Zealand, and there is no erasing or eradicating it. Disinformation, in its most insidious, liminal and porous forms, is contemptuous of sovereignty and borders. Every single internet connection at home or work, phone or PC, is a vector for harm, hate and violence. From multiplayer games, self-hosted group chats, private and decentralised cloud services, specific game console communities, augmented and virtual reality domains, the appropriation of emojis or memes to communicate hate, encrypted messaging, private groups and the dark web, disinformation actors and misinformation architects already have a plethora of platforms to infiltrate, and instigate socio-political unrest.

Official policies, laws and regulatory frameworks will never address the heterogenous assemblage of actors and platforms intent on undermining democracy, for two reasons. One, they have time on their side, and work towards intended outcomes years if not decades into the future using a combination of electoral, political, social and cultural means, over offline and online vectors. Two, the essential naïveté of social media companies, allowing till recently politicians to get away with inciting hate and violence results in, amongst other things, outdated and outmoded oversight, placing at risk communities who are often already marginalised, and have violence directed against them.

Laws and legislation are important, but very unlikely to address root causes and core motivations of growing disinformation concerts. What more can and should be done?

Corresponding with the principles laid out in the Global Network Initiative's (GNI) framework study on addressing digital harms and protecting human rights, first principles included in the Universal Declaration of Human Rights and the UN Guiding Principles for Business, leading social media companies are embracing a rights-based approach to governance, after years of a more laissez-faire approach. There is a

timely, rich and vital discussion that flows from this Silicon Valley pivot for domestic regulators and policy makers. For example, issues like responsibility, responsiveness, proportionality and transparency find renewed focus in regulatory conversations after the violence in Capitol Hill on 6 January 2021. Aotearoa, New Zealand however can and must delve deeper into disinformation's drivers. How can enclaves of resistance and immunity be crafted?

A good start would be to stop talking about online extremism or social media, and instead study the generation of violence and hate through broader ecological perspectives. Not unlike forestry or agriculture, factors influencing growth, pollination, yield, health and sustainability are invariably connected to context. What nourishes visible out-growth lies beneath what is often studied, or pared. The roots of discontent, often pre-dating online platforms by decades, are significant in the study of online content. Reciprocally, the vector, volume and velocity of digital content influences offline relationships and developments, especially around emotive issues, contested histories, and marginalised communities. Data visualisation, analytics, cognitive neuroscience and emergent research on cognition security are only as useful as securing those who can locate digital data in corporeal lives, recognising that what's encoded online is the algorithmic representation of complex, fluid, embodied realities. Deconstructing the digital requires the researcher to be rooted in local cultures, which in Aotearoa, New Zealand means the radical reintegration of Maori perspectives in regulatory and policy discourses around disinformation.

This perspective, congruent with my own experience and research including representations of violence and prosocial responses on social media in Sri Lanka and Aotearoa, New Zealand, turns on its head current approaches to countering extremism, largely based on enhanced or increased regulation, legal and legislative means. Recalling the Christchurch Commission Report's emphasis on social cohesion, we must imagine a more grounded, ecological and inter-disciplinary approach to research and response. Indigenising the inoculation against disinformation gains from harvesting the rich imagination, experience and insights of the Maori in Aotearoa, New Zealand. Through how they (who are us) understand identity, community, society, discourse, remediation and reparation, we can co-construct new socio-political structures that through equitable and democratic offline representation strengthens online responses to injury, incitement or invective. This radical dialogue, based on, amongst other things, active listening, rights, reciprocity and social justice, can be a constant, grounded inquiry that, combined with other disciplines, sets up a comprehensive response to disinformation's well-springs.

To end a policy brief on the value of offline relationships is perhaps counter-intuitive, but a necessary course-correction to technocratic approaches to a socio-political issue. Doctoral research, comparing social media in Aotearoa, New Zealand and Sri Lanka, supports the view that offline relationships, including political culture and the quality of journalism, significantly (and, at times, predominantly) influence online discourse. Our digital selves imagine a world as it can or should be, while our embodied selves negotiate the world as itis. This friction is essentially violent, and will always be so. Embracing this, enlightened socio-political and technological responses need to imagine stronger, more representative, endogenous and indigenous frameworks against threats to democracy in online and offline fora.

Why? Because *He waka eke noa.*

Video and text of presentation on Reimagining Extremism: Context, culture, community and country [17]

ICT4Peace Foundation's Special Advisor Sanjana Hattotuwa was invited by New Zealand's Department of Prime Minister and Cabinet to speak at He Whenua Taurikura, New Zealand's first annual hui (meaning a large gathering in Maori) on countering terrorism and violent extremism. The hui was held from 14-16 June 2021 in Christchurch. He Whenua Taurikura translates to 'a country at peace'. This presentation was delivered as part of the fourth panel at the hui, on day two.

Sanjana followed presentations by Jordan Carter from InternetNZ, Kate Hannah from Te Pūnaha Matatini and University of Auckland, Dr Nawab Osman from Facebook, Nick Pickles from Twitter, and Anjum Rahman from Inclusive Aotearoa Collective Tāhono and Islamic Women's Council of New Zealand. The panel was chaired by Paul Ash, head of the Christchurch Call.

Panel 4: Violent Extremism online: new directions in preventing radicalisation and violent extremism in the digital world

The internet is essential to modern life. The ability to connect individuals and share ideas across the world has delivered huge benefits. Connectivity is a force for good, but it has also empowered violent extremists who seek to inflict harm. This session will look at the role of the internet in violent extremism – from radicalisation and connecting extremist elements across the world, through to the sharing of violent and extremist content. The discussion will consider ways to prevent harm and keep people safe and secure, including efforts to address terrorist and 14 violent extremist content (TVEC) online. It will consider the role of multi-stakeholder collaboration, with a focus on the Christchurch Call, and the measures governments, industry, and civil society can each take on- and offline.

17 https://ict4peace.org/activities/presentation-on-reimagining-extremism-context-culture-community-and-country/

This session will discuss the following questions:

1. How do violent extremists (ab)use the online environment? What effect does this have on our safety and security?
2. What role do online environments – including social media and online algorithms – play in radicalisation? And in preventing radicalisation?
3. How do we make positive change in the online environment? What are the roles for government, industry and civil society?
4. This is a global problem. What international developments can we learn from in Aotearoa New Zealand? And what unique contribution can we make?

###

In seven slides, Sanjana articulated a vision for a radical revision in definitions of, discourse around and policymaking to address disinformation and violent extremism. He began by critically framing the chair's opening remarks, noting that geographic exclusion won't stop Aotearoa New Zealand from dealing with, at increasing pace, threats to its democratic integrity and institutions. He also noted that though recent reports of anti-Maori and minority online violence generated significant discussion and pushback, racism, violence against minorities and discrimination was not a new phenomenon in Aotearoa New Zealand. He also questioned the Chair's assertion that "no one group has the definitive voice" on the internet, noting that a critical gaze around privilege, power asymmetries including platform governance and policies by social media companies, the voice of G7 or OECD countries over the collective experiences of the Global South and factors including identity, language, location and gender, were central in biases inherent in information flows on the internet, and especially featured in social media.

With these initial comments, Sanjana introduced the central theme of his presentation, which was around an eco-systems or ecological perspective on disinformation, misinformation and violent extremism. Noting that ecosystems are heterogenous, endogenous and in Aotearoa New Zealand, indigenous, Sanjana flagged now approaching complex issues from a systems perspective aided in strengthening what binds and unites society, as much as the study of and responses against that which seeks to divide it, and seed a democratic deficit.

Slide 2 synthesised two-decades of activism, research and work in the domain of technology for peacebuilding, which Sanjana noted took him to five continents. Through

this lived experience, where embodied realities and the existential negotiation of violence, manifest through online vectors as well, Sanjana noted how early research from Sri Lanka led to early warnings around the instrumentalisation of social media platforms to incite violence and hate. Flagging the fluid and complex nature of disinformation landscapes, Sanjana noted how prolonged research – embracing big data, anthropological and ethnographic perspectives, grounded data to context – led to brains of disinformation researchers that were wired differently (using an image of trees around Lake Wanaka in winter). A patina of violence, experienced and studied, Sanjana noted, allowed researchers to see novel and enduring connections between online and offline content.

In slide 3, Sanjana presented a new way of looking at threats to democracy arising from hate innovation, autocratic expansion and authoritarian political entrepreneurs. Stressing that the discourse required a new vocabulary and imagination to fully capture the contours of, Sanjana used the analogy of surfing to explain how what entities like Aotearoa New Zealand's intelligence services and government were geared to respond to stochastic, episodic moments and short-term impacts, without adequate study of who producers were, intentionality, strategy and long-term entrenchment. The emphasis on the long-term strategy, and impacts, Sanjana averred was necessary since disinformation architects were working beyond timescales bound by the lifespan of a single PM, government, with influence operations that often lay beyond the directly or easily observable phenomena.

In slide 4, Sanjana recalled a point in the policy brief written for the hui,

> *Like a digital Novichok, the manner in which society sees itself, negotiates difference, communicates with each other, deals with the past, and envisions the future – and an individual's or community's place in it or ownership of it – can be corrupted through online content and social media platforms. Unlike a nerve agent however, which has an immediate and visible physiological impact, through influence operations conducted over time, the tone, timbre and thrust of divisive frames can become the foundations of political and social discourse. Sociologist Diane Vaughan called it "the normalisation of deviance" in relation to what caused the Challenger Space Shuttle disaster in 1986. Over time, individuals can come to accept a problem as a feature, instead of an aberration. The bad actors become those amongst us – our extended family, friends and neighbours – who come to believe in things we can no longer identify with, or subscribe to.*

Sanjana noted that those in Aotearoa New Zealand, used to a certain liberal-

democratic order, political culture, media landscape and way of life, may lack the imagination and thus the ability to pose critical questions around the nature of disinformation operations and strategies far more prevalent in the Global South. In a related point, Sanjana noted that while the Christchurch attacker was a foreigner, the instrumentalisation of media and offline vectors meant that over time, accelerated by issues related Covid-19 including vaccine hesitancy, those who spread misinformation and contributed to the evaporation if not evisceration of democracy, would be citizens of Aotearoa New Zealand. Like the pictured Craters of the Moon geothermal park in Rotorua, Sanjana noted that unrest in society could grow, without being noticed, if the questions asked around countering violent extremism weren't fit for purpose.

In the next slide, Sanjana used photos from Dunedin's Botanical Garden to present an ecological perspective of policymaking and regulation. This slide builds on discussions with Aotearoa New Zealand's Chief Censor and others from government around a radically new approach to policymaking, that eschewed the codification of laws and regulation as an end point or solution, and instead focussed on process, iteration and responsive all-of-government and all-of-society mechanisms as guardrails against violent extremism's seed and spread. Noting that disinformation, misinformation and extremism were socio-technological and socio-political issues, Sanjana stressed the need to move away from technocratic solutionism, highlighting the importance of rights and roots. The interplay of light and darkness, which trees grew where and how, and essentially, to approach the domain as the Department of Conservation would approach a park's ecological health, Sanjana noted, shifted the conversation, study and response.

Using murmuration as an example, Sanjana flagged now swarm dynamics could help explain and, in Aotearoa New Zealand, strengthen social cohesion, recommended by the Royal Commission Report on the Christchurch massacre. Noting that the princples of social cohesion (belonging, inclusion, participation, recognition and legitimacy) could be mapped on to the three fundamental principles governing murmuration (adhesion, cohesion and repulsion) Sanjana said that studying what attracted, retains and results in the rejection of violent extremism, at scale, could help better inform policymaking.

In the next slide, Sanjana called for a new imagination, flagging the work of Thomas Wright, who centuries before Hubble, presented for the first time the spiral shape of the Milky Way galaxy and spoke of the possibility of many solar systems.

In the final slide, against the background of a photo from Doubtful Sound and based

on the poet Browning's call to reach beyond what was immediately within one's grasp, Sanjana flagged how the response to terrorism, extremism and disinformation in Aotearoa New Zealand was fundamentally different to similar and contemporaneous discussions in other contexts and countries. Reaching into his experience of working on these issues in the Balkans, South Asia, Southeast Asia, and on both sides of the Atlantic, Sanjana emphasised the importance of meaningfully incorporating, as first principles, the Maori including definitions of community, country, culture, ground, identity, time, reparation, redress, the negotiation of difference and conflict transformation.

About ICT4Peace Foundation

ICT4Peace is a policy and action-oriented international Foundation. The purpose is to save lives and protect human dignity through Information and Communication Technology. Since 2003 ICT4Peace explores and champions the use of ICTs and new media for peaceful purposes, including for peace-building, crisis management and humanitarian operations. Since 2007 ICT4Peace promotes cybersecurity and a peaceful cyberspace through inter alia international negotiations with governments, international organizations, companies and non-state actors.

The ICT4Peace project was launched with the support of the Swiss Government in 2003 with a book, published by the UN ICT Task Force in 2005 on the theory and practice of ICT in the conflict cycle and peace building[18] and the approval of para 36 of the Tunis Commitment of the UN World Summit on the Information Society (WSIS in 2005.[19]

ICT4Peace on Twitter - www.twitter.com/ict4peace

ICT4Peace on Facebook - www.facebook.com/ict4peace

ICT4Peace official website: www.ict4peace.org

ICT4Peace additional publications: www.ict4peace.org/publications

ICT for peace foundation

---

18 https://ict4peace.org/wp-content/uploads/2019/08/ICT4Peace-2005-Information-and-Communication-Technology-for-Peace.pdf

19 Para *36. We value the potential of ICTs to promote peace and to prevent conflict which, inter alia, negatively affects achieving development goals. ICTs can be used for identifying conflict situations through early-warning systems preventing conflicts, promoting their peaceful resolution, supporting humanitarian action, including protection of civilians in armed conflicts, facilitating peacekeeping missions, and assisting post conflict peace-building and reconstruction.*