Check for updates

The Christchurch Call: insecurity, democracy and digital media - can it really counter online hate and extremism?

William James Hoverd[1] · **Leon Salter[2]** · **Kevin Veale[2]**

Abstract

The Christchurch Call was an international collaborative pledge between nation states and online service providers "to eliminate terrorist and violent extremist content online." In this article, we set out to provide an integrated cross-disciplinary analysis of the implications of the limitations of the Christchurch Call. We argue the existence of the Call helped change the conversation on the role played by online communication in hate, harassment and terrorism. However, the Christchurch Call is limited in its ability to counter online hate and extremism. Its current policy framework is most likely to produce messaging that shields social media platforms and other key figures from their existing responsibilities in producing insecurity on and offline. In particular, it does not address the wicked problem of how to understand the social, communal and individual dynamics when the online expression of free speech turns to hate, and in turn, violence—together with the role platforms and states play in permitting and intervening in such digital interactions.

Keywords Christchurch Call · Christchurch terror attack · Social media · Digital hate · FAANG · Harassment communities · Democracy · National security

Introduction

In the aftermath of the March 15th 2019 terror attacks on the Christchurch Al-Noor Mosque and Linwood Avenue Islamic Centre, the New Zealand (NZ) Government undertook four centralised regulatory responses. These included new restrictive firearms laws in combination with gun buy backs (Every-Palmer et al 2020), a *Royal Commission of Inquiry into the Attack on the Christchurch*

✉ William James Hoverd
w.hoverd@massey.ac.nz

1 Centre for Defence and Security Studies, College of Humanities and Social Sciences, Massey University, Wellington, New Zealand

2 School of English and Media Studies, College of Humanities and Social Sciences, Massey University, Wellington, New Zealand

Mosques, and the domestic ban of the attacker's online manifesto. Fourthly, and of particular relevance here, Prime Minister Jacinda Ardern initiated a global call to bring together social media company executives and heads of state in an attempt to control digital hate and extremism. This global initiative to eliminate online terrorist and violent extremist content is labelled *The Christchurch Call* (the Call).

As of August 2020, the Christchurch Call has not been rendered to significant academic scrutiny (Pandey 2020). This is due to a current focus on the immediate consequence and aftermath of the attack, and because quantifying the efficacy of the Christchurch Call is itself problematic. Study of the Call is affected by one of the persistent research challenges for studies of online hate, harm and extremism: the lack of primary data, with many studies reliant on secondary analysis (Schuurmann and Eijkman 2013; Berentson-Shaw et al. 2019). Empirical work has been emerging on white supremacist content, participation and recruitment online (Wong et al 2015; Scrivens et al 2019, 2020). However, the study of online hate is fraught with methodological challenges surrounding the systematic identification and collection of speech that can be labelled hate (MacAvaney et al. 2019). As a result, academic study of the Christchurch Call is reliant solely upon a critical analysis of the primary sources produced by official government discourse.

In the New Zealand context, there are three additional limitations that occur when discussing the sources around the Christchurch attacks. Following the New Zealand tradition of excluding publicity to the attacker (Walqhuist 2019), rendering him nameless, this article does not name the attacker except when quoting other's work. It employs the male pronoun or the term 'the attacker' instead. Likewise, due to the attacker's *The Great Replacement Manifesto* being banned by the Chief Censor in New Zealand (Christchurch Attacks Classification Information 2019), we can only refer to this primary source through secondary sources. Lastly, given the ongoing *Royal Commission of Inquiry into the Christchurch Mosque attacks*, we are unable to elicit additional primary data outside what is currently publicly provided by the NZ Government.

Given these primary data limitations, innovation is required to study the regulation of digital hate attempted by the Christchurch Call. Pandey has recently noted that the Call has an inherent tension between protecting against threat and safeguarding democratic freedom (Pandey 2020). Much of the commentary around the Call has focused on either practical/policy questions of countering terrorism (See Battersby and Ball 2019) and the regulation of social media platforms (Thompson 2019a), or social/cultural questions around racism (Herrara and Sabaratnam 2019), links to fascism (Sparrow 2019) and Islamophobia (Mirnajafi and Kate Barlow 2019). To date, little attention has been placed on how the policy and social domains interact. We argue that innovative, cross-disciplinary perspectives are essential in order to understand the implications of security policy for the encouragement/discouragement of online hate, which is, in essence, a cultural phenomenon.

In this article, we set out to provide a cross-disciplinary analysis of the implications of the Call from three domains of knowledge and practice:

(1) a national security perspective with a focus on online intelligence gathering and regulation,
(2) a digital media focus with a tradition of investigation into online harm, and lastly
(3) bringing that analysis into conversation with the critical study of liberal democracy.

Through this cross-disciplinary analysis, we set out to answer the following three questions:

(a) What are the national security goals and limitations of the Christchurch Call?
(b) What are the practical implications of applying the Christchurch Call to online hate speech and extremism?
(c) And, what are the implications of the Christchurch Call for the future of liberal democracy?

In combination, our cross-disciplinary analysis highlights the limitations of the Christchurch Call's national security objectives through a critical exploration of its attempted regulation of extremist content in the digital media space. The three voices approach the issue from three disciplinary perspectives. Our purpose is to open a multi-faceted debate around the Call's successes and identify its limitations. Finally, this analysis allows us to interrogate the consequences of such initiatives, and their natural tension with freedom of expression for the functions of liberal democracy.

We argue that the Christchurch terror attack demonstrates how notions of national security are tied into broader transnational, collective and social/cultural challenges around the polarisation of politics and the creation of insecurity within online spaces. An intervention along the lines of the Christchurch Call is a vital necessity to change a status quo which drives and profits from online insecurity (We broadly understand online insecurity to be a wide range of potential *risks*, *harms* and *threats* emerging from and connected to online spaces). However, the details of the current Call's focus on threat undercuts its goals in several key areas. Firstly, without broader nation state and transnational corporate engagement and proper oversight, the Call cannot be truly effective for national security purposes. Additionally, the Call seems unaware of the tensions around the social costs of regulation that would be required to achieve its aims. Secondly, it is hamstrung by the assumption that there is a clear separation between physical and online spaces, or between terrorism and normal harassment. The Call also does not engage with the structural factors fuelling online insecurity, including the ways that technology companies are either indifferent to insecurity or profit from it. As a result, the Call either ignores or misunderstands vital dynamics underlying online hate and violence, which means its impact will be limited. Lastly, the Call makes no mention of 'democracy,' which is unusual given that one of the aims of hate speech and terror attacks are the withdrawal of the targeted groups from public discourse. This leads to a loss of these marginalised voices from the public sphere. In addition, we suggest that the term 'Terrorist' as it is applied in the Christchurch Call risks becoming an empty signifier,

in favour of creating positive branding for the global digital (FAANG) platforms, who by appearing to work with nation states, discourage scrutiny of how their business models encourage online and offline hate. We argue that the Christchurch Call is a mechanism that could lead to positive change in a complex area. However, its current policy framework is most likely to produce messaging that shields social media platforms and other key figures from their existing responsibilities in preventing insecurity on and offline. We suggest in the conclusion, further cross-disciplinary research is required to refine and develop the goals of the Christchurch call to remedy the problems we identify in this paper.

The Christchurch terror attack and social media

The embedded usage of social media to facilitate the Christchurch terror attack was a prominent feature of the attacker's particular *modus operandi*. Prior to the attack, he allegedly announced his intentions on 8chan, and had earlier distributed his *The Great Replacement* manifesto[,12] widely across a variety of internet platforms. As the attack occurred, it was broadcast on Facebook's Live Stream service. An insight into the attacker's embedded social media extremist approach to violence can be found in his alleged post to 8Chan:

> Somewhere between 10 and 20 min before the first mosque was attacked, ******, logged on to the /pol/section of 8chan, an imageboard popular with the extreme right. As an anonymous user, ****** announced himself with a post entitled "*ahem*." It read: "Well lads, it's time to stop shitposting and time to make a real-life effort post. I will carry out and [sic] attack against the invaders, and will even live stream the attack via facebook." He then allegedly posted the link to his account (*******9), which was subsequently removed. "By the time you read this I should be going live." The post was also a farewell and indicated that he had been a frequent user of the platform. "I have provided links to my writings below, please do your part spreading my message, making memes and shitposting as you usually do. If I don't survive the attack, goodbye, godbless and I will see you all in Valhalla! (Macklin 2019).

In the attacker's alleged 8chan post, the author indicated their familiarity with social media, as well as the intent to use these platforms as a means to distribute their extremist views. They announced an intention to live record their violence to a global audience. The explicit intention is to spread these actions and views as widely as possible. Indeed, the social media spread of the attacker's message through the

---

[1] Both the manifesto and live streaming of the attacks has been deemed objectionable material by the New Zealand Office of Film and Literature Classification. As such, being subject to New Zealand law we have not accessed either of these two media in the preparation of this article. See https://www.classificationoffice.govt.nz/news/latest-news/christchurch-attacks-press-releases/ last accessed 11th May 2020.

[2] Based on the Great Replacement ideas of anti-immigration writer Frenchman Renaud Camus, https://www.great-replacement.com/ last accessed 11th May 2020.

manifesto and the live stream of the murders on Facebook was, for his purposes, extremely effective. Commentators suggest that some 4000 people watched the live video on Facebook and that some 1.5 million uploads of the video occurred in the subsequent 24 h (Thompson 2019a; Besley and Peters 2020). Thompson (2019a) notes that the live feed was also uploaded to 8chan and 4chan, and was distributed by a variety of prominent news media organisations across the world. The broad dissemination of the attacker's extremist content effectively demonstrates that the internet allows for the transmission, perpetuation and recycling of racist and unacceptable viewpoints (Besley and Peters 2020). Consequently, the ongoing distribution of the attacker's video and manifesto across social media was used to amplify (Macklin 2019) the effects of the terror attack, and functions as a potential call to arms for future copycat attacks.

The Christchurch Call: a national security perspective

Since the Facebook live streaming of the March 2019 Christchurch Mosque terror attack, one of the New Zealand Government's core national security responses was the global Christchurch Call initiative (Hoverd 2019, p. 22). The Call was initiated by NZ Prime Minister Jacinda Ardern who, in collaboration with French Prime Minister Emmanuel Macron, co-chaired a Christchurch Call to Action Summit in Paris on May 15th 2019. According to the New Zealand Government:

> The meeting aimed to see world leaders and CEOs of tech companies agree to a pledge called the 'Christchurch Call' to eliminate terrorist and violent extremist content online (Ardern 2019b).

Announcing the initiative on May 16th 2019, Ardern argued that:

> The March 15 attack was shocking in its use of social media as a tool in the act of terror and with the Christchurch Call we have taken a unique approach to solving this problem. (Ardern 2019a).

The Call is described as being collective in nature, asking Governments and online service providers to address and censure the promulgation of terrorist and extremist material online. Ardern stated that:

> The Christchurch Call is an action plan that commits government and tech companies to a range of measures, including developing tools to prevent the upload of terrorist and violent extremist content; countering the roots of violent extremism; increasing transparency around the removal and detection of content, and reviewing how companies' algorithms direct users to violent extremist content. (Ardern 2019a).

The Call's central goal is to counter the promulgation of violent, extremist content on social media platforms. Implicit in the logic behind the Call is the security argument made by the New Zealand Government when justifying the need for the summit: "It's critical that technology platforms like Facebook are not perverted as a tool

for terrorism, and instead become part of a global solution to countering extremism (Ardern 2019b).” The summit requested participants sign a three-part non-binding pledge to support its goals. By September 24th 2019, some 47 countries (excluding the United States who cited support for the summit but claimed to be constrained by the First Amendment) and eight online service providers had signed the Christchurch Call pledge.[3] The signatories committed to various actions depending on whether they are a nation state or a social media company (Call 2019).

This global engagement was sought because the Christchurch Call website argues that the internet can impinge upon “Collective security” (Call 2019). The Call acknowledges that violent social media content is available globally, but manifests locally, and in physical spaces within nation states. We argue that the Christchurch Terror attack has clearly shown how notions of national security are tied into broader transnational, cultural and social challenges around the polarisation of politics and the creation of insecurity within online spaces. We suggest that social media platforms provide a space for a variety of individuals to create and find communities where their feelings of insecurity can be expressed through processes such as shitposting,[4] and legitimate the sharing of extreme ideas with likeminded communities. In doing so, a process of community occurs that insulates an individual’s worldview through the ostracism of others, in this case Muslim immigrants (Herrara and Sabaratnam 2019; Mirnajafi and Kate Barlow 2019). This process, while not physically violent, falls into a category of hate and harm, which does not stay confined to online spaces.

For nation states, the Christchurch Call has opportunistic implications for developing the surveillance and regulation of social media by intelligence apparatus through increased cooperation with social media companies. For the surveillance mechanisms of a State, the question of how to generate open source intelligence from social media has been vexing (Trottier 2015). Social media intelligence has emerged as a developing opportunity for the overt and covert collection of information for security agencies (Wells and Gibson 2017; Bányász, 2018). However, the field has been dogged by a lack of technical expertise. Moreover, legislative frameworks are lacking that would allow for online intelligence collection, or provide the ability to use online data as court evidence for prosecution or the production of search warrants (Hassan 2019; Waddington 2019). The challenge for social media companies and national security agencies is to discern which online shitposts are best understood as simply mischievous speech acts, which are best understood as

---

[3] The initial signatories included Australia, Canada, European Commission, France, Germany, Indonesia, India, Ireland, Italy, Japan, Jordan, The Netherlands, New Zealand, Norway, Senegal, Spain, Sweden, United Kingdom, Amazon, Daily Motion, Facebook and Google.

[4] The term shitposting has evolved over time, and it has a long history. KnowYourMeme summarises it as “a range of user misbehaviours and rhetoric on forums and message boards that are intended to derail a conversation off-topic,” and it places the origins of the term as no later than 2007 (Unknown 2014). Shitposting is associated with low-effort posts and/or obnoxiousness. The daily dot.com’s critiques of Donald Trump supporters, see https://www.dailydot.com/layer8/trump-centipede-btfo-cuckold/ (last accessed 11th May 2020) summarise shitposting as “a troll, a deliberate provocation designed for maximum impact with minimum effort.”

harmful, and which represent potential physical threats in everyday life such as in the Christchurch attack. The question becomes, how should State intelligence apparatus appropriately, proportionally and legally discern threat and intent in online spaces? The attacker's alleged 8chan post clearly demonstrates this insecurity challenge. This post indicates that it is time to stop shitposting, where shitposting is simply chatter not backed by action or intent, and make a real-life violent effort. The enforcement issue for security agencies is to decide when to intervene. Aside from these enforcement issues, the broader question becomes how can democratic nation states' national security apparatus protect their populations against the potential incubation of such violent threats across social media, while still maintaining a free unregulated internet?

Ultimately, the effectiveness of the Christchurch Call for preventing the online promulgation of digital hate is questionable. The notion of 'whack-a-mole' is a known and continual challenge for security agencies (Andersen 2017), in that if you suppress a threat in one locale, it simply shifts to another. 8Chan continues to spread material on the Christchurch attack and other subsequent terror attacks. The security-oriented initiatives of the Christchurch Call can easily be circumvented. Inevitably, a variety of other unregulated social media spaces have simply become alternative mediums for the expression of digital hate.

The Christchurch Call is a well-meaning start to address such (in)security challenges, but fails to ultimately address the problem of whack-a-mole. Unless it is accompanied by broader nation state and transnational corporate cooperation, proper oversight and legislation, it cannot be truly effective for national security purposes. And even if it were to be effective, the ever-expanding amount of surveillance infrastructure and regulation of internet freedom that this would entail would be a heavy social cost. Moreover, there are significant implications for how we think about digital media content and liberal democracy, if we are to consider such a possibility. We turn to discuss these now.

The Christchurch Call: a digital media perspective

The structure of the Christchurch Call to Action Summit offers positive opportunities for fostering change in the landscape of social media and online spaces, where the underlying business models normalise harassment and extremism. However, the Christchurch Call is limited by its focus on explicit terrorist acts committed in physical space, as if there was a clear separation both between online and offline contexts, or between the harm caused by terrorism and online harassment. The reality is that there is no meaningful distinction between online and offline spaces, and those who commit terrorist acts are merely the most active members of whole communities dedicated to harming other people. Although the Christchurch Call's structure presents a promising opportunity, without accounting for these dynamics its impacts are likely to be limited.

For example, the Call summit began with the statement that the white supremacist terrorist attack in Christchurch was 'unprecedented,' ('Christchurch Call' 2019) but such a claim ignores a significant amount of history and context. Christchurch

was not the first time that social media platforms have been implicated in terrorism: however this is the first time that a terrorist attack in a 'western' country was broadcast via the internet, but Facebook has been a significant factor in the genocide of Rohingya Muslims in Myanmar, for example, as covered in the Frontline documentary "The Facebook Dilemma" (Jacoby 2018). Additionally, a study called "Fanning the Flames of Hate: Social Media and Hate Crime" by Karsten Müller and Carlo Schwarz demonstrated a link between Facebook use and violence against refugees in Germany (Müller and Schwarz 2018; Taub and Fisher 2018). Social media has been connected to acts of terrorism and broader social abuses for a long time, with explicit acts of terrorism being merely the tip of the iceberg produced by a vastly more substantial broad base of problem content and communities—and these events are ignored by the current framing of the Christchurch Call.

Christy Dena introduced a useful model for understanding dynamics within online hate groups while writing about Alternate Reality Games (ARGs): she uses the term 'tiers' to describe how ARG communities stratify around different levels of engagement (Dena 2008, pp. 42–43, 2009, pp. 239–258). In broad strokes, the members of the primary tier are the most active members of an ARG, who bring in new material; the secondary tier fits that material together; and the tertiary tier forms an audience that engages with the output of the other tiers. People move between tiers as their levels/types of engagement fluctuate.

Harassment communities also exhibit all of these dynamics, except that the context of tiering adapts to a situation where the goal is committing concrete harm to someone's ability to live their life (Veale Forthcoming). The challenge of the 'game' comes from overcoming any resistance provided by the people being terrorised as they try to protect themselves and those close to them. The tertiary tier functions almost exactly as it would for a normal ARG, and is made up of people who are following the activities of the harassment community by supporting them without participating themselves. The secondary tier of harassment communities seek opportunities to capitalise upon and promote particular achievements made by those in the primary tier. Those achievements encapsulate a diverse set of activities because of how wildly diverse the activities of the primary tier itself is – something also true of normal ARGs.

The individuals within the primary tier do incredible amounts of labour to forward the cause of the harassment community, regardless of what mode that labour happens to be. They are most likely to be the people who can be personally identified for their contributions: part of the motivation for the labour is to achieve social capital (or infamy) within the harassment community (Butt and Apperley 2016; Veale 2013). Examples would include people like Benjamin "Bendilin" Daniel, who created the "Beat Up Anita Sarkeesian" videogame as part of the harassment campaign targeting her in 2012, which was then popularised and circulated by members of the secondary tier, drawing it to the attention of the third 'audience' tier (Klee 2014; Sarkeesian 2012, 2014a, b). Similarly, the people who produced and sold videogames glorifying the terrorist attacks in Christchurch (which were then banned as objectionable material by the NZ Office of Film & Literature Classification) would also qualify as working within the primary tier (O'Connor 2019b; Tait 2019). For harassment campaigns more generally, the

individuals who uncover personal information through deep research and circulate it to the community are in the primary tier, as are the people who directly harass targets either in-person, through physical proximity, or through media like phones (Campbell 2017; Dewey 2016; McKibben 2016).

All of the tiers of a harassment campaign are united in their focus on causing concrete harm (mental, emotional, physical and material) to the people they target: consequently, there is no meaningful distinction between online and offline spaces (Veale Forthcoming). The intent of harassment communities is to take the damage to the real world. For example, "Google bombs" are campaigns where a coordinated mob of attackers post vast amounts of material connecting their target to a crime they want to frame them for (Citron 2014, pp. 69–72). If enough articles are posted claiming that someone is a terrorist, or a paedophile, etc., then otherwise neutral web-searches for that person's name will increasingly return articles claiming that they're a criminal threat (Lum 2015; Pless 2015; Wilson 2016). Often, the next step is to contact the workplaces of the people the community is targeting in order to question how the company can justify employing a depraved criminal. If the employer does a web-search to verify these claims—claims magnified greatly via repetition through the harassment community—then it will look like the accusations have substance. Harassment campaigns using Google bombs have resulted in several known cases where people have been fired, and there are doubtless many other less high-profile examples where individuals have had their livelihoods threatened or destroyed. Another example of harassment campaigns seeking real-world harm is through SWATTING—false emergency calls to encourage police to raid a target's home, which often suggest the target is armed and a danger to other civilians—a crime which has resulted in more than one death (Cross 2015; O'Connor 2019a; Sinders 2015). Part of the effectiveness of harassment campaigns is that they suppress resistance and ethical dissent. If anyone speaks up in support of someone being targeted by a harassment campaign, they will also be targeted—alongside their friends and loved ones, and those of the initial target. The harm is delivered from an online space, but does not stay there.

Within the primary tier of harassment campaigns are a subset of individuals willing to both threaten credible physical violence and then to carry it out (Robertson 2014; Sarkeesian 2014a, b), such as this example from Brianna Wu:

> I got home from a movie with my husband, and someone had sent me pictures of standing right behind me in the movie theater, just to say, hey, I know where you live. (Cornish 2019).

This example shows the interaction between online and offline harassment, in that someone within the online community was motivated to stalk Wu, and to share photographic evidence with the community. The community then ensured she knew she was being followed. Another chilling example of this dynamic is Elliot Rodger, who murdered six people in 2014 after posting a sexist 'incel' manifesto online on 4chan. Rodger has been praised as a 'saint' by some online harassment communities—with further killers directly claiming him as inspiration for their attacks on women (BBC 2018a, b).

Kiwi Farms is a community that has driven more than one person to suicide (Fogel 2018; 'lightninggrrl' 2016; Pless 2016). Both 8chan and Kiwi Farms have been linked to multiple mass killings that were celebrated in their communities (Hankes 2018; Neiwert 2015). It is in this light, that we have to understand the Christchurch attacker referring to the livestream of his massacre as an "effort post" on 8Chan—in contrast to a low-effort "shitpost" (Rowe 2019). His livestream was an exemplar first-tier attempt to court the social capital and approval of existing white supremacist harassment communities online. As such, it is a claim to infamy that most within the community can only aspire to—with the horrifying threat that some will be *inspired* to do so.

Currently the Christchurch Call is focused on the most explicit expressions of terrorist intent expressed online, and thus it ignores the underlying disease of harassment communities in favour of its most obvious symptoms: exemplar first-tier individuals operating in physical spaces. Its current framework misses communities like Kiwi Farms or the act of Google Bombing entirely. Additionally, it fails to address the growth, propagation and recruitment of harassment communities that, at their extremes, encourage more people to commit mass murders for notoriety, social capital and their pet cause.

The Call also doesn't engage with the structural dimensions to escalation within harassment communities, and the role played by social media platforms and internet infrastructure companies in perpetuating and expanding the secondary and tertiary tiers of harassment communities. For example, Rebecca Lewis' work charts how Youtube's recommendation algorithm helps foster and connect a spreading network of neo-nazis who prefer the more neutral branding of "alt-right." Moreover, the monetisation of Youtube content and advertising allows them to make money from hate and spread their influence in a way that Youtube itself also profits from (Lewis 2018). There is a fundamentally capitalist motivation for the escalation and perpetuation of harassment, through being rewarded by Youtube's algorithm.

The current aims and targets of the Christchurch Call either ignores or misunderstands the vital tiering dynamics underlying online hate and violence, and the role social media platforms play in profiting from and encouraging the communities from which primary-tier exemplar violence comes. As a result, its impact going forward will be limited, despite the fact that the international, multi-level collective action offered by the Call presents a model by which positive changes could otherwise be made.

Where's democracy? The Christchurch Call and the turn to terror

This section focuses on the absence of the subject of democracy from the Christchurch Call document, and the implications this may have for the exclusion of marginalised groups from public discourse due to digital hate. Following on from the previous section, we continue the assertion that the activities of those labelled *terrorists* cannot be clearly separated from other forms of digital hate, which, while not necessarily physically violent, causes symbolic violence, leading to significant psychological damage or even suicide. Our specific focus here is the capacity or otherwise of

the digital public sphere to include the voices of "alternative, marginalized, or otherwise oppressed groups" (Dahlberg 2011, p. 863). In reference to Muslim communities, they have the added impediment of being associated with terrorism. Hence, we argue that in the Christchurch Call document, and in a subsequent speech made by Facebook COO Sheryl Sandberg at the UN, there is a danger that *terrorist* becomes an empty signifier, enabling the positioning of the FAANG (Facebook, Apple, Amazon, Netflix, Google) as part of an *us* which excludes the very victims of the Mosque shootings and discourages further scrutiny of the relationship between their profit model and the proliferation of online hate.

The word democracy is entirely absent from the Christchurch Call document. This absence seems stark given the substantial focus in the aftermath of Brexit and the Trump election, by national governments, platforms, and several academic disciplines, on the question of the increasingly negative impacts of platform algorithms on the normal functioning of Western liberal democracies (see for example Deb et al. 2017; Hicks 2018; Howard et al. 2019; Persily 2017). These authors highlight the risks to democracy of an increasingly polarised and hateful digital public sphere, something that we argue cannot be solved by technical or policy means alone (see also Kuehn and Salter 2020).

As Thompson (2019a, p. 5) has argued, "terrorist" content is only the "tip of the iceberg". Removing it treats only the very worst symptoms, not the majority of symptoms, let alone the cause. Further, extremist, terrorist or radicalising content *is* hateful content, with the aim of forging an in-group community through shocking, racist and/or discriminatory depictions of an out-group (Bangstad 2014; Lewis 2018). Empirical studies have demonstrated that hateful online content disproportionately affects people of colour (Gardiner et al. 2016), women (Chen 2018; Edstrom 2016) and Muslims (Bangstad 2014), and that it discourages these groups from engaging and participating in public discourse, for fear of further intensified hate (Bangstad 2014; Chen 2018; Edstrom 2016; Jakubowicz 2017).

From the 1990s through to the early 2010s, the internet was heralded as holding the potential for the realisation of a unifying and inclusive Habermasian public sphere (Habermas 1991), which would be the locus of "rational communication and public opinion formation" (Dahlberg 2011, p. 859). With its enablement of disembodied discourse, and ease of access, digital media was thought to have the capacity to widen national-level policy conversations to include "previously marginalized individuals and communities" (College of St George 2018, p. 1). This discourse can still be detected within various government 'e-democracy' drives, including New Zealand's. A recent report entitled *How Digital Can Support Participation in Government* (Government Information Services 2018) stresses the importance of including marginalised groups such as Māori, as well as designing digital tools to enable "two-way engagement and deliberative discussion" (20). And the first line of the Christchurch Call document frames the internet as "a powerful tool to…enhance social inclusiveness".

We argue that to have any hope of realising these aims of enhancing and widening democratic participation, the issue of how the algorithms of the FAANG valorise hate (both economically and culturally) must be addressed. Further, we place caution on an over-liberal use of the term "terrorist", given that signifier has

an ongoing pejorative association with Muslims since 9/11 (for example through the 'War on Terror'). This association has been used to justify very high levels of state surveillance on Muslim communities in the US and UK. It has also contributed to their feelings of alienation from mainstream society (Blackwood et al. 2013). And as outlined by Carpentier (2008) in his analysis of political discourse during the Iraq War, the signifier 'terror/terrorism/terrorist' creates simplifying friend/enemy distinctions, "defining the horizon of our thought and excluding other discourses" (30).

We therefore argue that the signifier "terrorist" can become what Ernesto Laclau (1996) terms an "empty signifier". Empty signifiers by their nature are dynamic. When a hegemonic alliance (such as between Facebook and national governments) uses a term like "terrorist" to signify its unity, it can contribute to broader societal changes in its meaning. This hegemonic alliance relies on a symbolic enemy, which incarnates the values represented as opposite to those of the in-group alliance. This was evidenced when Facebook COO Sheryl Sandberg spoke with Jacinda Ardern at a follow-up Christchurch Call meeting, following their twin addresses to the UN in New York in September 2019. Less than a week earlier, Facebook had launched their own policy changes, independent of the requirements of the Call (Facebook 2019). Facebook's PR machine has been forced to go on the front-foot in the last three years, as its profit model has become implicated in a series of scandals, including Cambridge Analytica, the Trump election, and a genocide in Myanmar. Focusing on the terrorist signifier provides Facebook with a decoy, relieving them of some of the substantial scrutiny that's been rightly directed their way. While the Christchurch Call document does make reference to "the operation of algorithms," there is a danger that debates have subsequently been allowed to shift onto FAANG's terms: the problem has been relabelled as a few bad eggs taking advantage of Facebook's drive to give everyone voice and enable a free marketplace of ideas [see also a recent speech by Mark Zuckerberg at Georgetown University (IANS 2019)]. Referencing her speech at the UN General Assembly, Sandberg claimed afterwards:

> There was not one country represented in that room that had not been touched by terrorism and violent extremism. And the terrorist's goals are very clear: they aim to silence, they aim to stand against the values that we hold so dear, they aim to have people live in fear. And our goals are exactly the opposite – we want people to have voice, we want people to live with humanity and dignity. We want people to speak for tolerance and against hatred. (quoted in Cheng 2019).

"The terrorist" is here operating as an empty signifier, a common-enemy which allows Sandberg to position Facebook as part of *us*, sharing "the values that we hold so dear". These values are presumed to be the ability to "have voice", which, along with "connectivity," produces more data, and more profits for social media corporations. Therefore, rather than challenge a profit model which reproduces (in)security through its algorithms, we can see that the Call is becoming a vehicle for the sustainment of that status quo, as well as the dissemination of Silicon Valley ideologies which equate security and dignity with connectivity.

Conclusion

The Christchurch Call has helped change conversations about the current role played by the interactions of social dynamics and algorithms to produce digital hate, harassment and terrorism. Massive international public pressure brought a coalition of global governments, major technology companies and social media companies to a table in an unprecedented way. This form of structure, by opening a space for the debate of "the issue of social media regulation" (Thompson 2019b, p. 84) presents opportunities going forward. However, the specific policy frameworks discussed by the Christchurch Call have substantial holes that undercut its stated aims, which become very visible when explored from a cross-disciplinary perspective.

Firstly, without broader nation state and transnational corporate buy in and proper oversight, it cannot be truly effective for national security purposes, and it seems unaware of the tensions around the social and bureaucratic costs of the regulation that would be required to achieve its aims. The Christchurch Call is a well-meaning start to address (in)security challenges, but it ultimately fails to address the problem of whack-a-mole, where perpetrators of violence simply shift locale as a response to state interventions. For intelligence and prosecutorial purposes, at this stage the Call offers no practical way forward for nation states. Secondly, the Christchurch Call has a problematic focus on explicit terrorist acts committed in physical space, as if there was a clear separation both between online and offline contexts, or between the harm caused by terrorism and online harassment. The current aims and targets of the Christchurch Call either ignores or misunderstands the vital tiering dynamics within harassment communities and the algorithmic profiteering that encourages the communities from which primary-tier exemplar violence comes. Lastly, the Call makes no mention of 'democracy.' The term 'Terrorist' as applied in the Christchurch Call risks becoming an empty signifier that alienates the people targeted in the attack in favour of creating positive branding for massive corporations that discourages scrutiny of how their business models encourage online and offline hate. The current agreement ignores the role that social media companies still play in fostering harassment communities from which first-tier exemplars (terrorists) might emerge.

In essence, the Christchurch Call is a mechanism that could lead to positive change in a complex area. However, its current policy framework is most likely to produce messaging that shields social media platforms and other key figures from their existing responsibilities in producing insecurity on and offline.

We suggest further research which develops the original intent and principles of the Christchurch Call to address these limitations, because the current framework creates an illusion of progress and action while masking deeper insecurity. If nation states and social media companies genuinely wish to be seen as acting against the kinds of terrorist action and communication that the Christchurch Call stands against, then closer attention needs to focus on the capitalistic model that sustains harassment communities. In addition, the wicked problem is how to understand the social, communal and individual dynamics when the online expression of free speech turns to hate, and in turn, violence—together with the role platforms and states play in permitting and intervening in such digital interactions.

Compliance with ethical standards

Conflicts of interest There are no applicable conflicts of interests or competing interests.

References

Andersen L (2017) The MOLE and the mallet: Islamic state and al-qaeda in the "thirty years" war' in the middle east. Connections 16(1):7–24

Arden J (2019b) NZ and France seek to end use of social media for acts of terrorism. 23rd April 2019 https://www.beehive.govt.nz/release/nz-and-france-seek-end-use-social-media-acts-terrorism. Accessed 24 April 2020.

Arden J (2019a) Christchurch Call to eliminate terrorist and violent extremist online content adopted. May 16th 2019, https://www.beehive.govt.nz/release/christchurch-call-eliminate-terrorist-and-violent-extremist-online-content-adopted. Accessed 24th April 2020.

Bangstad S (2014) The weight of words: the freedom of expression debate in Norway. Race Class 55(4):8–28. https://doi.org/10.1177/0306396813519928

Bányász P (2018) Social media and terrorism, AARMS: academic and applied research in military science. Budapest 17(3):47–62

Battersby J, Ball R (2019) Christchurch in the context of New Zealand terrorism and right wing extremism. J Polic Intell Count Terror 14(3):191–207

BBC (2018a) 'Toronto Suspect Praised "incel" Killer'. BBC News, 25 April 2018, sec. US & Canada. https://www.bbc.com/news/world-us-canada-43883052. Accessed 23 April 2020

BBC (2018b) 'How Rampage Killer Became Misogynist "Hero"'. BBC News, 26 April 2018, sec. US & Canada. https://www.bbc.com/news/world-us-canada-43892189. Accessed 23 April 2020

Berentson-Shaw J, Elliott M, Kuehn K, Salter LA (2019) Digital threats to democracy. The Workshop. https://www.digitaldemocracy.nz/. Accessed 3 Sept 2020

Besley T, Peters MA (2020) Terrorism, trauma, tolerance: bearing witness to white supremacist attack on Muslims in Christchurch. N Zealand Educ Philos Theory 52(2):109–119. https://doi.org/10.1080/00131857.2019.1602891

Blackwood L, Hopkins N, Reicher S (2013) Turning the analytic gaze on "us": the role of authorities in the alienation of minorities. Eur Psychol 18(4):245–252. https://doi.org/10.1027/1016-9040/a000151

Butt MR, Apperley T (2016) 'Vivian James: the politics of #Gamergate's Avatar'. In From the 1st Joint International Conference of DiGRA and FDG. The School of Arts, Media and Computer Games, Abertay University, Dundee

Campbell C (2017) Anita Sarkeesian's Astounding "Garbage Human" Moment. Polygon. 27 June 2017. https://www.polygon.com/features/2017/6/27/15880582/anita-sarkeesian-garbage-human-vidcon-interview. Accessed 23 April 2020

Carpentier N (2008) The construction of the self and the enemy in the 2003 Iraq war. In Spinoy NCE (ed) Discourse theory and cultural analysis: Media, arts and literature. Hampton Press, Cresskill

Chen H-T (2018) Spiral of silence on social media and the moderating role of disagreement and publicness in the network: analyzing expressive and withdrawal behaviors. New Media Soc 20(10):3917–3936. https://doi.org/10.1177/1461444818763384

Cheng D (2019) Christchurch Call update: social media giants join forces to fight extremism. https://www.nzherald.co.nz/nz/news/article.cfm?c_id=1&objectid=12269942. Accessed 23 April 2020

Christchurch Attacks Classification Information (2019) https://www.classificationoffice.govt.nz/news/latest-news/christchurch-attacks-press-releases/. Accessed 22 April 2020

Christchurch Call (2019) The Christchurch Call to action: to eliminate terrorist and violent extremist content online. https://www.christchurchcall.com/christchurch-call.pdf. Accessed 23 April 2020

Citron D (2014) Hate crimes in cyberspace. Harvard University Press, Cambridge

College of St George (2018) Democracy in a post-truth information age: background paper. Windsor: a St George's house consultation. https://www.stgeorgeshouse.org/wp-content/uploads/2017/10/Background-Paper.pdf

Cornish A (2019) 'How gamergate became a template for malicious action online'. All Things Considered (NPR). https://www.npr.org/2019/08/30/756034720/how-gamergate-became-a-template-for-malicious-action-online. Accessed 23 April 2020

Cross K (2015) '"Things have happened in the past week": on Doxing, Swatting, and 8chan'. Feministing. https://feministing.com/2015/01/16/things-have-happened-in-the-past-week-on-doxing-swatting-and-8chan/. Accessed 23 April 2020

Dahlberg L (2011) Re-constructing digital democracy: an outline of four 'positions.' New Media Soc. 13(6):855–872. https://doi.org/10.1177/1461444810389569

Deb A, Donohue S, Glaisyer T (2017) Is social media a threat to democracy?. https://www.omidyargroup.com/wp-content/uploads/2017/10/Social-Media-and-Democracy-October-5-2017.pdf. Accessed 23 April 2020

Dena C (2008) Emerging participatory culture practices: player-created tiers in alternate reality games. Convergence 14(1):41–57. https://doi.org/10.1177/1354856507084418

Dena C (2009) 'Transmedia practice: theorising the practice of expressing a fictional world across distinct media and environments'. PhD, Sydney: University of Sydney. https://www.christydena.com/phd/. Accessed 23 April 2020

Dewey C (2016) 'This horrifying and newly trendy online-harassment tactic is ruining careers'. The Washington Post. https://www.washingtonpost.com/news/the-intersect/wp/2016/04/11/this-horrifying-and-newly-trendy-online-harassment-tactic-is-ruining-careers/. Accessed 23 April 2020

Edstrom M (2016) The trolls disappear in the light: Swedish experiences of mediated sexualised hate speech in the aftermath of Behring Breivik. Int J Crime Just Soc Democracy 5(2):96–106

Every-Palmer S, Cunningham R, Jenkins M, Bell E (2020) The Christchurch mosque shooting, the media, and subsequent gun control reform in New Zealand: a descriptive analysis. Psychiatry Psychol Law. https://doi.org/10.1080/13218719.2020.1770635

Facebook (2019) A further update on the New Zealand terrorist attack. https://about.fb.com/news/2019/03/technical-update-on-newzealand/. Accessed 16 Oct 2020

Fogel S (2018) 'Video game developer dies after setting herself on fire'. Variety (blog). https://variety.com/2018/gaming/news/chloe-sagal-death-1202858068/. Accessed 23 April 2020

Gardiner B, Mansfield M, Anderson I, Holder J, Louter D, Ulmanu M (2016) The dark side of Guardian comments. The Guardian. https://www.theguardian.com/technology/2016/apr/12/the-dark-side-of-guardian-comments

Government Information Services. (2018). How digital can support participation in government. The Department of Internal Affairs, New Zealand Government, Wellington. https://www.digital.govt.nz/standards-and-guidance/engagement/online-engagement/research-how-digital-can-support-participation-in-government/

Habermas J (1991) The structural transformation of the public sphere: an inquiry into a category of bourgeois society. MIT Press, Cambridge

Hankes K (2018) 'Evidence of New Mexico school shooter's involvement in the racist "alt-right" is overwhelming'. Southern Poverty Law Center. https://www.splcenter.org/hatewatch/2018/02/08/evidence-new-mexico-school-shooter%E2%80%99s-involvement-racist-alt-right-overwhelming. Accessed 23 April 2020

Hassan N (2019) Gathering evidence from OSINT sources: digital forensics basics. Apress, Berkeley

Herrera A, Sabaratnam A (2019) After Christchurch standing up to racism [online]. Advocate 26(1):36–37

Hicks T (2018) Defending and recovering american election systems. Brown J World Affairs 24(2):97–108

Hoverd W (2019) The changing New Zealand National Security Environment: new threats, new structures, and new research. Natl Secur J 1:17–34

Howard PN, Ganesh B, Liotsiou D, Kelly J, François C (2019) The IRA, social media and political polarization in the United States, 2012–2018. Project on Computational Propaganda, Oxford. https://comprop.oii.ox.ac.uk/research/ira-political-polarization/. Accessed 23 April 2020

IANS (2019) Facebook now 'fifth estate', take it seriously: Mark Zuckerberg. Gulf News. https://gulfnews.com/technology/facebook-now-fifth-estate-take-it-seriously-mark-zuckerberg-1.1571384459114#. Accessed 23 April 2020

Jacoby J (2018) The Facebook Dilemma. Frontline. PBS. https://www.pbs.org/wgbh/frontline/film/facebook-dilemma/. Accessed 23 April 2020

Jakubowicz A (2017) Alt_Right White Lite: trolling, hate speech and cyber racism on social media. Cosmopolit Civil Soc Interdiscipl J 9(3):41

Klee M (2014) 'Creator of "Beat Up Anita Sarkeesian" Says #gamergate is anti-harassment'. The Daily Dot. https://www.dailydot.com/parsec/creator-beat-up-anita-sarkeesians-says-gamergate-is-anti-harassment/. Accessed 23 April 2020

Kuehn KM, Salter LA (2020) Assessing digital threats to democracy, and workable solutions: a review of the recent literature. Int J Commun 14:22

Laclau E (1996) Emancipation(s). Verso, London

Lewis R (2018) 'Alternative influence: broadcasting the reactionary right on Youtube'. Data & Society. https://datasociety.net/output/alternative-influence/. Accessed 23 April 2020

'lightninggrrl' (2016) 'I Am Being Stalked and Harassed by Kiwi Farms and SA'. Wrong Planet. https://wrongplanet.net/forums/viewtopic.php?t=308671. Accessed 23 April 2020

Lum Z (2015) 'Veerender Jubbal, Sikh-Canadian Journalist, Wrongly ID'd As Paris Terror Suspect'. Huffington Post. https://www.huffingtonpost.ca/2015/11/16/veerender-jubbal_n_8577520.html. Accessed 23 April 2020

MacAvaney S, Yao H-R, Yang E, Russell K, Goharian N, Frieder O (2019) Hate speech detection: challenges and solutions. PLoS ONE 14(8):e0221152. https://doi.org/10.1371/journal.pone.0221152

Macklin G (2019) The Christchurch terror attacks, livestream terror in the viral video age. Combating Terrorism Center, Westpoint. July 2019, Volume 12, Issue 6.

McKibben B (2016) 'embarrassing photos of me, thanks to my right-wing stalkers'. The New York Times. https://www.nytimes.com/2016/08/07/opinion/sunday/embarrassing-photos-of-me-thanks-to-my-right-wing-stalkers.html?smid=tw-share&_r=1. Accessed 23 April 2020

Mirnajafi Z, Barlow FK (2019) Beyond tokenistic solidarity in the wake of the Christchurch terrorist attacks: insights from psychology, and practical suggestions for action. N Zeal J Psychol 48(1):47–55

Müller K, Schwarz C (2018) Fanning the flames of hate: social media and hate crime. SSRN. https://doi.org/10.2139/ssrn.3082972

Neiwert D (2015) 'Illinois woman with neo-nazi leanings charged in Canadian mass murder plot'. Southern Poverty Law Center. https://www.splcenter.org/hatewatch/2015/02/18/illinois-woman-neo-nazi-leanings-charged-canadian-mass-murder-plot. Accessed 23 April 2020

O'Conno A (2019a) 'Court gives 20 years in prison for swatting death spark by CoD match'. Rock, Paper, Shotgun (blog). https://www.rockpapershotgun.com/2019/04/01/20-years-prison-sentence-for-swatting/. Accessed 23 April 2020

O'Connor, A (2019b) 'New Zealand banned a mass shooting game as a "terrorist publication"'. Rock, Paper, Shotgun (blog). https://www.rockpapershotgun.com/2019/11/01/new-zealand-banned-a-mass-shooting-game-as-a-terrorist-publication/. Accessed 23 April 2020

Pandey P (2020) 'One year since the Christchurch Call to action: a review,' ORF Issue Brief No. 389, August 2020, Observer Research Foundation. https://www.orfonline.org/wp-content/uploads/2020/08/ORF_IssueBrief_389_Christchurch.pdf. Accessed 28th Aug 2020

Persily N (2017) The 2016 US election: can democracy survive the internet? J Democracy 28(2):63–76

Pless M (2015) '5 reasons i stand with Sarah Nyberg'. Internet Famous Angry Men. https://idledillettante.com/2015/09/11/5-reasons-i-stand-with-sarah-nyberg/ Accessed 23 April 2020

Pless M (2016) 'Kiwi farms, the web's biggest stalker community'. New York Magazine. https://nymag.com/selectall/2016/07/kiwi-farms-the-webs-biggest-community-of-stalkers.html. Accessed 23 April 2020

Robertson A (2014) 'Trolls drive Anita Sarkeesian out of her house to prove misogyny doesn't exist'. The Verge. https://www.theverge.com/2014/8/27/6075179/anita-sarkeesian-says-she-was-driven-out-of-house-by-threats. Accessed 23 April 2020

Rowe D (2019) 'The online cesspits where hate found a home'. The Spinoff (blog). https://thespinoff.co.nz/media/19-03-2019/the-online-cesspits-where-hate-found-a-home/. Accessed 23 April 2020

Sarkeesian, A (2012) 'Anita Sarkeesian at TEDxWomen 2012'. TEDxTalks. https://www.youtube.com/watch?v=GZAxwsg9J9Q

Sarkeesian A (2014) 'I usually don't share the really scary stuff. But it's important for folks to know how bad it gets [trigger warning]'. Tweet. @femfreq (blog). https://twitter.com/femfreq/status/50471 8160902492160/photo/1. Accessed 23 April 2020

Sarkeesian A (2014) 'Anita Sarkeesian speaking at XOXO conference'. Feminist Frequency. https://femin istfrequency.com/video/anita-sarkeesian-speaking-at-xoxo-conference/. Accessed 23 April 2020

Schuurmann B, Eijkman Q (2013) Moving terrorism research forward: the crucial role of primary sources. International centre for counter terrorism. ICCT background paper. The Hague

Scrivens R, Davies G, Frank R (2020) Measuring the evolution of radical right-wing posting behaviors online. Deviant Behav 41:216–232

Scrivens R, Gaudette T, Davies G, Frank R (2019) Searching for extremist content online using the dark crawler and sentiment analysis. In: Deflem A, Silva D (ed) Methods of criminology and criminal justice research (sociology of crime, law and deviance, vol. 24). Emerald Publishing Limited, pp 179–194

Sinders C (2015) 'that time the internet sent a SWAT team to my mom's house'. Narrative.Ly. https://narrative.ly/that-time-the-internet-sent-a-swat-team-to-my-moms-house/. Accessed April 23 2020

Sparrow J (2019) Facists among us. Online hate and the Christchurch Massacre. Scribe Publications, Victoria

Tait M (2019) 'Two terrorist publications banned'. NZ Office of Film & Literature Classification. https://www.classificationoffice.govt.nz/news/latest-news/two-terrorist-publications-banned. Accessed 23 April 2020

Taub A, Fisher M (2018) 'Facebook fueled anti-refugee attacks in Germany, new research suggests'. The New York Times. https://www.nytimes.com/2018/08/21/world/europe/facebook-refugee-attacks-germany.html. Accessed 23 April 2020

Trottier D (2015) Open source intelligence, social media and law enforcement: visions, constraints and critiques. Eur J Cult Stud 18(4–5):530–547

Thompson P (2019a) Beware of geeks bearing gifts: assessing the regulatory response to the Christchurch Call. Polit Econ Commun 7(1):83–104

Thompson P (2019b) Beyond the Christchurch Call: a scoping framework for developing regulation & policy for social media & digital intermediaries. University of Wellington, Victoria

Unknown (2014) 'Shitposting'. Know your meme. 3 March 2014. https://web.archive.org/save/https://knowyourmeme.com/memes/shitposting. Accessed 4 May 2020

Veale K (2013) Capital, dialogue, and community engagement—'my little pony: friendship is magic' understood as an alternate reality game. Transform Works Cult. https://doi.org/10.3983/twc.2013.0510.Accessed23rdApril

Veale K (Forthcoming) Gaming the dynamics of online harassment. Palgrave Macmillan, London

Waddington D (2019) Introduction: the police and social media. In: Akhgar B, Bayerl P, Leventakis G (eds) Social media strategy in policing security informatics and law enforcement. Springer, New York

Walqhuist C (2019) Ardern says she will never speak name of Christchurch suspect. The Guardian Online. https://www.theguardian.com/world/2019/mar/19/new-zealand-shooting-ardern-says-she-will-never-speak-suspects-name. Accessed 23 April 2020

Wells D, Gibson H (2017) OSINT form a UK perspective: consideration form the law enforcement and military domains. In: Proceedings Estonian Academy of Security Sciences 16: From Research to Security Union. Estonian Academy of Security Sciences, pp 84–113.

Wilson J (2016) 'Game developer group doesn't let nintendo off the hook for Firing Alison Rapp'. Venture Beat. https://venturebeat.com/2016/04/05/game-developer-group-doesnt-let-nintendo-off-the-hook-for-firing-alison-rapp/. Accessed 23 April 2020

Wong M, Frank R, Allsup R (2015) The supremacy of online white supremacists: an analysis of online discussions by white supremacists. Inf Commun Technol Law 24:41–73