Collaborative Engagement with the Social Media Companies – Wrap up

Overview

The Christchurch terror attack on 15 March 2019 was a significant point in New Zealand's history. The premeditated attack on two Christchurch Mosques took the lives of 51 New Zealanders and severely impacted many more. It was a direct assault on the country's cherished ideals of multiculturalism.

By capturing an act of terror live on social media and by using the internet as a tool to boost exposure to the killings, the gunman ensured his hateful agenda was maximally amplified. The social media platforms were also used to share the gunman's manifesto. The actual horrific attack was live-streamed for a full 17 minutes before action began to try and remove it.

The video of the Christchurch murders was online for nearly an hour before New Zealand Police flagged it with Facebook. The company's algorithms had failed to recognise the nature of the damaging content. Copies of the live-stream went viral despite attempts to shut diffusion down. The video, in its various forms, reached millions of viewers and can still be found online today. The safeguards in place were drastically inadequate for preventing global distribution of the attack.

The terrorist attack came at a time of escalating levels of investor discomfort relating to the social media platforms. Investors had long-term concerns around, for example, poor corporate governance practices and dual class share structures leading to heavily skewed voting control. More recently, concerns had focused on issues such as the Cambridge Analytica scandal, allegations of electoral manipulation and the use of social media platforms to spread of misinformation and hate speech.

For the Guardians of New Zealand Superannuation (Guardians), the widespread dissemination of the Christchurch murders crossed a line compelling us to take action. In our view, Facebook, Alphabet and Twitter had betrayed their users trust, breached their duty of care and severely damaged their social licence to operate. The Guardians' team decided to rally like-minded investors to join together and engage these three main social media companies with a single focus: to strengthen controls to prevent the livestreaming and dissemination of objectionable content.

Starting locally, we gained incredible support from New Zealand investors. As the collaboration grew, invitations were expanded to international peers. We recognised that to successfully get our message to the leaders of these large multinational corporations, we needed the power of a large collaboration, speaking with a united voice, on an issue that represented both a moral imperative and a business case.

Ultimately, 105 global investors representing approximately NZD$13.5 trillion AUM joined the Social Media Collaborative Engagement over the ensuing months.

The business case for engagement

The issue of objectionable content being disseminated through social media platforms has severe and wide-reaching implications for investors, companies and the general public.

Technology stocks are a significant part of many global indices and as ESG risks have crystallised, we have seen consequences for global investment portfolios. Further, there are many additional risks for the broader technology sector. For example, the decline of consumer trust, litigation risk including anti-trust, General Data Protection Regulation compliance, regulatory risk, reputation risk and cyberattack risks.

These risks are compounded by the serious societal consequences of allowing objectionable material to be shared across social media platforms. We have seen this play out in various forms across the world.

Action

The first lever of action was to speak out publicly on our intention to engage the identified social media companies – Facebook, Alphabet and Twitter – on this issue. This was a deliberate step-change from our usual approach of confidential engagement with investee companies and was a clear stand against what we felt was a serious breach of accountability by the social media companies.

Second, the Guardians sought to build a global investor collaboration with the intention of using a large amount of assets under management to leverage a unified voice. One of the key factors in enabling such a large group to form was by establishing a single, clear objective. This helped to break down the barriers for those investors who had not joined a collaboration previously and ensured there was a clear and uniting goal for the engagement.

Third, as the global collaboration grew, we announced formal support of the Christchurch Call, a joint initiative by the governments of New Zealand and France which outlines collective, voluntary commitments from Governments and online service providers intended to address the issue of terrorist and violent extremist content online.

The group also created and distributed an investor resource for shareholders not part of the Collaboration who sought to engage on the same issue. This ensured the social media companies were hearing the same message from a wide range of investors, signalling the amplified importance of the issue to investors.

Engagement letters were sent to the Chairs of the Boards of each of the three companies on behalf of the Collaboration, and the Guardians secured engagement meetings with each company to discuss their responses to the Christchurch attack.

The social media companies assured the Collaboration they were making changes to strengthen controls. However, none of the companies agreed for a Board member to meet with the collaboration. With such a large group of influential investors behind this agenda, we did not feel there was enough commitment from the companies to let the matter settle.

Enhancing pressure over time

As the first anniversary of the Christchurch terror attacks approached, the investor group felt it was not being heard at an appropriately senior level. The Collaboration had become frustrated with progress and the inability to discuss concerns directly with the different Boards.

To compound this frustration, two more terror attacks (in Germany and Thailand) had been live-streamed across the social media platforms reflecting that the platforms were still open to abuse.

In response, the collaboration published an Open Letter, distributed via global press, calling for:

- Clear lines of governance and accountability to ensure social media platforms cannot be used to promote objectionable content like the live-streaming and dissemination of the Christchurch terrorist attack; and
- Sufficient resources dedicated to combatting the live-streaming and spread of objectionable material across the platforms.

Other tools considered were raising a shareholder resolution or campaigning to vote against a particular Director at the companies' annual meetings. However these were ruled out because of voting control issues due to multi-class voting structures limiting shareholders' ability to meaningfully influence via these key investor tools.

However, the Guardians did signal their voting intent to the Collaboration[1] and social media companies in advance of their Annual General Meetings and exercised its voting rights as follows:

- By withholding votes or voting against directors who were up for re-election and had not carried out their responsibilities as they relate to the live-streaming and dissemination of content; and
- Supporting shareholder resolutions, which in some way drove progress towards meeting the objective of its engagement.

Working to overcoming key challenges

The Collaboration has held a number of meetings with key executives and has continued, unsuccessfully, to seek meetings with Board Directors. We have used a range of tactics to try to overcome this barrier including: using the whole power of the collaboration to request a meeting, using a subset of influential investors to engage, offering the Guardians' CEO Matt Whineray to meet with the Board and using the influence of a top 10 shareholder, Northern Trust Asset Management, an active participant of the collaboration to reinforce our message. The continued lack of access to Boards remains a significant source of frustration for the investor group.

Results

In late 2020, Facebook informed us that they had strengthened the Audit and Risk Oversight Committee charter to explicitly include a focus on the sharing of content that violate its policies. It also included a commitment not just to monitor and mitigate such abuse, but also to prevent it. This notable improvement is directly attributable to this engagement and a real strengthening of governance and accountability for the Board on this issue. It puts the company on the front foot in working towards prevention of the issue rather than just fire-fighting inherent problems.

It is worth noting that since the Christchurch attacks, the platforms have all moved to strengthen controls to prevent the live streaming and distribution of objectional content. However, it is a difficult job for investors to assess if these changes are appropriate for the scale of the problem. Therefore, the Collaboration commissioned some external research to help better assess the matter.

The research was undertaken by a New Zealand based independent consultancy and think tank called Brainbox Institute. Brainbox specialises in issues at the intersection of technology, politics, law and policy. We chose Brainbox because of their deep understanding of the technical aspects of how the platforms operate and interlink with society. The findings of their review[2] on the technology changes concluded that:

- The measures introduced by the platforms have a high likelihood of significantly mitigating the scale of dissemination of future objectionable content.
- However, the platforms are highly unlikely to absolutely prevent the spread of objectionable content of another similar type of event because once new content has been uploaded, there is an unavoidable time delay before it can be accurately classified as objectionable. The platforms cannot eliminate this time-gap entirely.
- Until content has been classified as objectionable, there is effectively no measure the platforms can introduce that could entirely prevent user exposure to objectionable content.
- The Platforms have made and continue to make reasonable efforts to reduce the spread of objectionable content.

---

[1] *The Guardians did not issue voting guidance and made it very clear that every organisation should make their own voting decisions. There was no intention to act together with anyone on any votes or to form a shareholder "group" (as defined in Federal securities laws) for the purpose of acquiring, holding, voting or disposing of securities of Facebook, Alphabet, Twitter or any other issuer.*

[2] *The Brainbox Institute Report reflects its own views and not necessarily those of the Collaboration or any of its members. The Collaboration takes no responsibility for the accuracy of the report.*

- Restrictions on live streaming would not prevent the upload of non-livestreamed copies of objectionable content.
- The platforms are well-place to rapidly triage potential objectionable content and they have implemented mechanisms to quickly intervene in such cases: in fact, they can intervene much faster than any government body could. This means they will always have a significant role to play.
- The most effective means of dealing with a live crisis and the sharing of related objectionable content attacks comes from cross-platform collaboration efforts. In the specific case of violent online terror related content, use of a shared hash database as administered by the Global Internet Forum to Counter Terrorism (GIFCT) and the development of a 'Crisis Response Protocol' as part of the Christchurch Call, are key mechanisms to enable suppression of objectionable content that stems from a live crisis. These measures provide a procedure for platforms to rapidly coordinate, identify and classify objectionable content, and add information to a shared database so duplicate uploads can be quickly identified. However, they are not failproof and do have some limitations.
- Automation and the use of algorithmic systems are inevitable and necessary for the platforms to moderate content successfully given the volume of information that crosses the platforms. Human moderators will always be needed because of the risk of inaccuracies in classification by automation.

Brainbox also recommended the following areas or further improvement:

- It is generally sensible to assume that more funding and resources directed at continued efforts to reduce the classification time-delay will be needed.
- Investors should advocate for continued improvement and transparency on measures to prevent proliferation of objectionable content.
- All measures to find and prevent the spread of objectionable content have trade-offs between the human rights of those exposed to the objectionable content and those using the platforms to share content, whether objectionable or not. We need protection from abuse by those with intent to use the platforms maliciously, including those with decision making powers, but, fundamentally, we also need the ability to freely express our views and share material of importance with society. The grey area - where legal content is harmful - is a very complex area to navigate and any restrictions put in place must be constantly assessed for balance and refinement. This would be best enhanced through inviting independent scrutiny and assessment.

We take some reassurance from Brainbox's findings that the measures introduced by the platforms have a high likelihood of significantly mitigating the scale of future objectionable content. However, the findings do not exonerate the companies from their ongoing duties to prevent objectionable content making it on to their platforms and being seen by innocent users of their platforms.

In this particular case of objectionable content relating to the Christchurch terror attack, it was the first time someone had so meticulously planned the dissemination of his agenda using social media. We take heed of Brainbox's finding that the platforms are highly unlikely to absolutely prevent the spread of objectionable content of another similar type of event. We also note that the changes made by the companies lessened the amount of content shared for similar events that have occurred since then (the 9 October 2019 terrorist attack in Halle, Germany and the 20 May 2020 terrorist attack in Glendale, Arizona).

We realise that the success or failure of the social media companies in moderating content and preventing abuse is likely to determine whether users stay on the platforms or move towards alternatives. Therefore, we expect the social media companies to avidly continue to take efforts to reduce the classification time-delay, remain focused on this issue and continually evolve crucial safeguards to prevent against abuse.

Continued improvement in this area is fundamental to the basic viability of the platform businesses and also to their ability to respond to a crisis.

We believe the companies are only on the start of their journey. They must keep this issue elevated as a core focus of the Executive and Board, with considerable resourcing and reporting on progress between boards and investors.

Regulation

The core focus of this engagement was to improve corporate conduct of the social media companies. However, the group of investors is also very aware that if the platforms are perceived as unable or unwilling to effectively moderate user-submitted content, then regulation by countries will likely ensue. In fact, State regulation is already emerging in some key jurisdictions. The regulatory trend, broadly speaking, is toward the consistent expansion of the categories of content that, increasingly, need to be controlled. The trend began several decades ago with regulation intended to control narrow and specific classes of content (i.e., Child Sexual Abuse Material), and now decades later, regulation is aimed at controlling more broad and vague classes of content – content that may not be illegal but can definitely be deemed harmful to society. We asked Brainbox to look at some of the emerging regulation and assess the pros and cons of the different approaches to try to provide perspective on what 'good' regulation looks like. These are ultimately matters for lawmakers in different jurisdictions to assess and involve balancing a wide range of complex and sometimes competing policy considerations. This report was intended to contribute to the public discussion in this area and does not necessarily represent all the investors views in all aspects.

The investors were mindful that this task is particularly complex involving a deep understanding of intricate nuances between the prevention of exposure to objectionable content, the human rights impact of livestreaming or dissemination on victims of atrocities, or of increasing the likelihood of similar events occurring, the protection of human rights such as the rights to, for example, free expression, free speech, free association and privacy and the potential, intentionally or not, that regulation can limit human rights, have perverse consequences or limit important information reaching society.

Brainbox themselves share the concerns expressed by human rights organisations about aspects of the current regulatory trajectory. Of the regulation pieces assessed, Brainbox were of the view that the EU Digital Services Act seems to be the piece of regulation with the most support from human rights bodies with a focus on measures to enhance transparency and auditability of platform content moderation systems and processes. Their report noted widespread opposition from a range of stakeholders towards Australia's Abhorrent Violent Material Amendments which requires the expeditious removal of abhorrent violent material with penalties for failure to remove it including fines of up to 10% of annual group turnover for corporations and imprisonment for up to 3 years for individual employees that fail to remove or refer content. Brainbox were of the view that opposition to this Act was justified because it could cause the companies to over-react and overly restrict content as a result.

When you lift up above the specific details, the heart of the problem goes to where accountability lies between platform users, platform owners and governments. It is a complex and grey area that is poorly defined and vastly wide ranging. Every situation of abuse or the spread of objectionable content across platforms is different. Every set of contextual circumstances is different. Yet a common set of policies designed by the companies, applies.

Australia developed its regulation because the companies hadn't moved fast or far enough in terms of taking accountability for objectionable content on their platforms. The EU Digital Services Act proposal is designed around a principled approach based on transparency and the resulting accountability it drives. It would standardise approaches to reporting on how content is being moderated, create a right of review and appeal measure for users to appeal against content moderation decisions and therefore requires platforms to explain how they made decisions and according to what factors. It would need some time to become effective from a behavioural change perspective. When comparing the two sets of legislation, you

have a carrot and stick approach versus a more principled approach. Both have merits and both have downsides.

Brainbox's research was able draw out a summary of robust legislative mechanisms that investors can look for and advocate for as regulation in this area emerges. These included:

- Developing regulation related to content moderation should be anchored in the language and law of human rights which contain within them acceptable and proportionate balancing between rights and freedoms.
- Good legislation will set rules and standards that are as clear as possible to distinguish between what kind of content is permitted and what kind of content is not permitted.
- Any decision that applies the law must be capable of review and appeal by a legal body, such as a court or tribunal.
- There must be a demonstrable connection between the kind of conduct being restricted by regulation and the kind of harm that is alleged to result.
- Legislation should require and foster transparency about what content moderation actions are being taken and why. These transparency requirements should be imposed on both states and platforms.
- Legislation should not unjustifiably limit individual privacy, including by requiring platforms to report users to governments based upon what they are saying or doing online.
- Legislation that imposes massive financial penalties is likely to influence platforms to take a more conservative course of action to limit their risk, including to over-remove content rather than risk a fine, which is likely to have disproportionate effects on freedom of speech or expression.
- The strongest case for regulation relates to the area of transparency and auditability of content moderation systems.

What next?

We are now 2.5 years after the atrocities that occurred in Christchurch. This tragedy will always be with us and we will never forget those who lost their lives and the pain and suffering caused to their families and friends.

As a group of investors who have sought change with these three companies, we take some reassurance from Brainbox's findings that the measures introduced by the platforms have a high likelihood of significantly mitigating the scale of future objectionable content. However, we also know that the platforms are highly unlikely to absolutely prevent the spread of objectionable content of another similar type of event. They must keep this issue elevated as a core focus of the Executive and Board, with considerable resourcing and open and honest reporting on progress between boards and investors. Therefore, we wind this engagement up with the message to Facebook, Twitter and Alphabet that we expect them to avidly continue to take efforts to reduce the classification time-delay, remain focused on this issue and evolve crucial safeguards to prevent against abuse.

We also raise our dissatisfaction that the companies have continued to decline our requests for a meeting with a Board member. We would like the message to reach the Board that we have sought to work with them to solve some of these key challenges, not against them. We see the successful management of these issues as critical to the long-term success of the companies. We do not expect companies to pursue profit at all costs to society and we expect them to carry out their duty of care with absolute resolve.

And finally, we note that the issue of content moderation is becoming one of the defining legal and socio-political issues of our time. It deserves its own body of specialist expertise stretching across a range of academia, law and policy. We urge the companies to open up their platforms to allow independent scrutiny of policies and related decisions and actions We hope that the engagement undertaken as part of this collaboration and research undertaken by Brainbox adds a useful and thoughtful perspective to this live debate.

Special thanks

A special thanks goes to:

- The New Zealand government owned investors who supported The Guardians with leading this initiative from day one. These were the Accident Compensation Corporation (ACC) of New Zealand, the Government Superannuation Fund Authority, Kiwi Wealth and the Board of Trustees of the National Provident Fund.
- Neuberger Berman and Northern Trust Asset Management for their unwavering support for the engagement, their help with getting meetings with the companies, their advice and counsel and their funding for the independent research commissioned.
- Aviva Investors for support in leading engagement meetings with Twitter.
- LGPS Central for their explicit support and contribution to the Engagement Resource for Investors.
- Trusted Advisors of this engagement: Aberdeen Standard, AMP Capital, Aviva Investors, BMO Global Asset Management, Church of England Pension Board, Federated Hermes EOS, Hesta, LGPS Central, Neuberger Berman, Nest Corporation, Ninety One, Northern Trust Asset Management and Robeco.
- Brainbox Institute for their dedication and involvement in this complex and live debate.
- All the 105 participating signatories to the collaboration:

Leaders Group (NZ Crown-owned investors)
NZ Super Fund (NZSF)
Accident Compensation Corporation (ACC)
Government Superannuation Fund (GSF)
National Provident Fund (NPF)
Kiwi Wealth (KW)

Participants – New Zealand
AMP Financial Services
ANZ New Zealand Investments
Apostle Funds Management
ASB
BNZ
Booster
Fisher Funds
Foundation North
Generate Investment Management
Harbour Asset Management
H.R.L. Morrison & Co Limited
Investment Services Group (Devon Funds, JMI Wealth, Select Wealth and Clarity Funds)
IWIinvestor
JBWere NZ
MAS
Milford Asset Management
Mint Asset Management
MyFiduciary Limited
Ngāti Awa Group Holdings Limited
NZ Funds
Pathfinder
PIE Funds/JUNO KiwiSaver Scheme
Public Trust
Rata Foundation
Salt Funds Management
Simplicity

Smartshares
Tauhara North No2 Trust
Trust Investments Management Limited
Trust Waikato
Westpac / BT Funds Management

Participants – International

Aberdeen Standard Investments
Adrian Dominican Sisters, Portfolio Advisory Board
AMP Capital (NZ and International)
AP1
AP2
AP3
AP4
AQR Capital Management
Australian Ethical
Aviva Investors
Axa Investment Managers
Bayerische Versorgungskammer
BMO Global Asset Management
Bon Secours Mercy Health
Brunel Pension Partnership
Cadmos Peace Investment Fund
Caisse de dépôt et placement du Québec
Christian Brothers Investment Services
Church of England Pensions Board
Church Commissioners
Common Interests Financial
Congregation of St. Joseph
Coöperatie DELA
Daughters of Charity, Province of St. Louise
Dignity Health
Domini Impact Investments
ECO Advisors
First Sentier Investors
Greater Manchester Pension Fund
Hermes EOS
Hermes Investment Management
HESTA
Hexavest
Irish Life Investment Managers
HSBC Global Asset Management
Legal & General Investment Management
LGPS Central
LG Super
Local Authority Pension Fund Forum (LAPFF)
Media Super
Mercer Global (including NZ)
Mercy Investment Services, Inc.
Merseyside Pension Fund
Mirova
The National Employment Savings Trust (NEST)
NEI Investments
Neuberger Berman
Newton Investment Management
Ninety One

Nomura Asset Management
Northern Trust Asset Management
Northwest Coalition for Responsible Investment
OPTrust
Oregon State Treasury
Pantheon Ventures
Providence St. Joseph Health
River and Mercantile
Robeco
RPMI Railpen
U Ethical Investors
USS Investment Management
Utilities of Australia Pty Ltd
VFMC
VicSuper
West Midlands Pension Fund
West Yorkshire Pension Fund