



The Christchurch Principles



ACKNOWLEDGEMENTS

Our contribution to the Paris Peace Forum is possible thanks to the generous support of Mr. Cliff Cook and family.

This project was made possible thanks to support from our collaborators The Workshop and AUT.

Cover image credit: CHRISTCHURCH CITY COUNCIL.

Auckland University of Technology is partnering with The Helen Clark Foundation while the Foundation becomes established. This aligns with the University's commitment to research for the public good, and public communication of research. The University supports the Foundation's aim to fund and facilitate research on a range of topics; however its support is not an endorsement of any individual item of work, output or statement made by the Foundation's staff, directors or patron.

This paper is covered by the Creative Commons Attribution License 4.0 International. When reproducing any part of this report, including tables, full attribution must be given to the report author.

ABOUT THE HELEN CLARK FOUNDATION	05
INTRODUCTION	06
THE CHRISTCHURCH PRINCIPLES	08
THEORETICAL FRAMEWORK FOR THE CHRISTCHURCH PRINCIPLES	10
APPENDIX	14
BIBLIOGRAPHY	21

The Helen Clark Foundation is an independent public policy think tank based in Auckland, New Zealand, at the Auckland University of Technology. It is funded by members and donations. We advocate for ideas and encourage debate, we do not campaign for political parties or candidates. Launched in March 2019, the foundation issues research and discussion papers on a broad range of economic, social and environmental issues.



OUR PHILOSOPHY

New problems confront our society and our environment, both in New Zealand and internationally. Unacceptable levels of inequality persist. Women's interests remain underrepresented. Through new technology we are more connected than ever, yet loneliness is increasing, and civic engagement is declining. Environmental neglect continues despite greater awareness. We aim to address these issues in a manner consistent with the values of former New Zealand Prime Minister Helen Clark, who serves as our patron.

OUR PURPOSE

The Foundation publishes research that aims to contribute to a more just, sustainable and peaceful society. Our goal is to gather, interpret and communicate evidence in order to both diagnose the problems we face and propose new solutions to tackle them. We welcome your support, please see our website helenclark.foundation for more information about getting involved.

On 15 March 2019, a gunman opened fire in two mosques in Christchurch, New Zealand, killing 51 people and injuring 50 more. During his attack at the Al Noor mosque, where the majority of victims were killed, the alleged perpetrator livestreamed his actions directly on Facebook via a helmet-mounted camera.

As later confirmed by Facebook, the livestreaming of the Christchurch terrorist attack did not activate its monitoring mechanisms and it was not until a user alerted Facebook to the video – 29 minutes after livestreaming of the attack started and 12 mins after it ended – that it became aware of the issue. By that point approximately 4,000 people had already viewed the video. It was then widely shared on Facebook, quickly replicated and shared on other platforms, including YouTube and Twitter and appeared on several news media outlets. Prior to the attack the gunman published a wide-ranging manifesto outlining a white supremacist viewpoint as justification for his actions. The manifesto remains accessible online.

Within the first 24 hours of the terrorist attacks, Facebook removed more than 1.5 million uploads of the video.

Facebook's delay in disabling the livestream video of the attack on its platform and quickly preventing the further uploading and dissemination of the video has thrown a spotlight on the capacity and willingness of social media platforms to rapidly and effectively respond to terrorist and harmful content online, both in New Zealand and globally.

Many questions are now being asked about why Facebook was so slow to act, what more could social media companies have done and what decisive action needs to be taken to restrict the livestreaming of extremist and violent content in the future and hold social media companies to account. Questions have also arisen over the enabling conditions within online platforms that create self-validating echo chambers which contribute to isolation and radicalisation, and how social media companies, governments, and civil society can respond.

Our project seeks to develop a set of principles, inspired by human rights and democratic theory, which social media companies, civil society and government can draw on when deciding what internet governance should look like. If given the chance to scale up our project, we would use this opportunity to consult more broadly both domestically in New Zealand and internationally, advocate for change, and develop practical recommendations for implementation. This would allow us to further refine our principles and ground them in the needs of those most impacted by current policy settings.

We believe that the public sphere should be a place of equal participation. As the internet, and social media especially, becomes an increasingly influential public space, these digital media have greater implications for political equality, both online and offline. In some regards, the rise of social media has created new ways for people to participate in public life, to overcome long-standing disadvantages. In other regards, however, it may pose threats, especially through the flourishing of harmful online content. Where equality of participation is impeded, a correction or redress is required. This often will involve finding an appropriate balance between the right to free expression and other rights, including the right to participate in public affairs.

We see the Christchurch Principles as complementary to efforts taken by the New Zealand government as part of the ‘Christchurch Call’. The Christchurch Principles support and sit alongside the Christchurch Call, a commitment by Governments and tech companies to eliminate violence and extremist content online. The Call was initiated by New Zealand Prime Minister, Jacinda Ardern, and French President, Emmanuel Macron who brought together Heads of State and Government and leaders from the tech sector to adopt the Christchurch Call on May 15, 2019.

This government-led initiative is limited to violent and extremist content, and while it is understandable for policy makers to focus on these areas where agreement is most likely to

be achieved, we see these Principles to be more ambitious and applicable to the broader concept of harmful content, in order to change the perception of what feels possible.

The Principles also seek to build on the platform of compassion that characterised the response to the attacks from the people of Christchurch and New Zealand, supported by the global community, and to ensure positive change can arise from the horror of the March 15 attacks.

Because the Christchurch attacks are just the tip of the iceberg. Violent, extremist content continues to spread; terrorist radicalisation continues to be incubated online; and democracies worldwide are struggling to respond effectively to ‘fake news’. And the challenges continue to evolve. New technologies and an increasingly interconnected world have the potential to improve human lives, foster peace, and advance development, but, if left unchecked, also have the potential to sow the seeds of discord and hate, and undermine the bonds of trust and reciprocity that democracy is founded upon. As such, the Christchurch Principles seek to strengthen existing democracies in the face of these challenges, and to encourage democratic engagement in the way that government, civil society and businesses exercise their responsibilities in the digital era.

THE CHRISTCHURCH PRINCIPLES

1. The principle of equal participation:

A well-functioning democracy is one where people within a political community have the opportunity to participate as equals in public life. When harmful online content impedes that parity of participation, remedy is required a responsibility which falls variously on states, businesses, and civil society.

2. The duty to protect:

States have a duty to protect human rights from violations that occur on- and off-line. Because democratic institutions are also vital to upholding rights in a number of jurisdictions, states also have a corresponding responsibility to protect democratic norms and practices. Protective actions by the State must be assessed in light of potential infringements of other rights or democratic principles.

3. The responsibility to respect:

Businesses have responsibilities to respect the rights of persons. These responsibilities apply directly to businesses' impact on people's rights, but also extend to their capacity to influence the wider social, political and economic context, which may have positive or negative effects on democratic norms and practices. Democratic institutions ought to be respected by businesses, because these enable rights to flourish – by giving effect to certain civil and political rights, and by creating favourable conditions for human rights to thrive, including social and economic rights. This includes norms such as trustworthiness, transparency, accountability, inclusivity, privacy, and reciprocity.

4. The responsibility to remedy:

When a person's rights have been violated, or a person is unable to exercise their rights, or the principle of equal participation is violated because democratic norms and practices are degraded, then there are obligations to remedy. These obligations are to be allocated among states, businesses and civil society actors according to responsibility, capacity and efficacy. It is noted that civil society actors have a special role in remedying harms from online content, because these actors are not encumbered by the same power dynamics as states and businesses, and hence the ethical issues this poses. Accordingly, the responsibility to remedy for states and businesses may involve enhancing the capacity of civil society actors.

5. The principle of structural change:

By taking a proactive stance to remedy the negative impacts of digital technologies, it is necessary to take a systems approach and to identify interventions at the structural level; for example, through governance structures, regulation to restore transparency, accountability, fair competition, and genuinely participatory and representative multi-stakeholder processes.

6. The duty of care:

States have a duty of care toward the impacts of regulatory policies, and businesses have a duty of care toward the consequences of releasing their products. A duty of care requires taking an evidence-based approach, by gathering evidence when it is lacking, undertaking a risk analysis, and monitoring the online and offline effects of new regulations and products to guide adaptive management.

7. The principle of democratic means:

Democratic ends can only be sustained through democratic means, so it is incumbent upon states, businesses and civil society organisations to integrate democratic practices into their own structures. This involves inclusivity, transparency and reciprocity.

8. The principle of decentralisation:

Power ought to be decentralised in the digital realm, as it is in the political realm, if it is to be consistent with democratic outcomes. This can be achieved through a range of policies: shareholder democracy for social media companies; inclusivity measures for companies and civil society organisations; antitrust measures by states; the creation of councils, charters and forums which enable better representation in decision making about social media policy.

9. The principle of inclusivity:

Inclusivity of diverse voices is a key governance principle in any context, but it is especially important for new digital technologies where the marginalisation and disadvantaging of certain groups – online and offline – is at stake. The unique perspectives of marginalised groups must be recognised, which includes the perspectives of indigenous peoples, women, LGBTQ+, non-citizens and recent migrants, geographic and demographic populations with comparatively poor access to the internet, and members of minority religions and cultures.

10. The principle of communicative action:

Democracy requires more than freedom of expression; it requires effective communication which enjoys public trust by being sincere, honest, reliable, intelligible, relevant and competent. If the balance between trustworthy and untrustworthy communication is overly skewed toward the latter, then democracies will struggle to deliver on their promise of public decisions that enjoy genuine legitimacy and reflect the informed will of the people. States, businesses and civil society actors all have responsibilities to create the conditions for trustworthy communication to flourish, through regulatory and non-regulatory means.

Ultimately, we need to answer the question – who is to be in charge? – in a way that works for us, as a public and as individuals, and returns us to the promise of democratic space that the internet originally offered.

THEORETICAL FRAMEWORK FOR THE CHRISTCHURCH PRINCIPLES

The Christchurch Principles propose a set of regulative principles for reducing the impacts of harmful online content. These principles are designed to uphold human rights and democratic institutions by informing the behaviour of states, businesses (especially digital technology companies), and civil society. The Christchurch Principles recognise that new digital technologies have had many positive effects on rights and democracy, by creating new tools that people can use to exercise and defend these institutions. However, this does not diminish the need to address the negative effects of harmful online content.

Harm is defined in terms of (1) rights violations, and (2) threats and obfuscations that discourage a person from exercising their rights. As examples of the latter, if a person chooses not to participate in the political process because of threats of violent repercussion, or because of misdirection through disinformation, then this is considered a harm. This is because, whether rights are being violated or unrealised, a person's needs and interests are not being met.

Underlying this view of harm, it is recognised that online content has a substantive relationship to offline behaviour and well-being. This is true both for the victims and perpetrators of harm. For victims of online harm, this may not only impact their capacity to share opinions on online platforms, but also offline in public and private spaces, because the effects of threats and discrimination may come to infuse their real-world life. For the perpetrators, harmful online

content can be a contributing factor in a person's radicalisation, which can motivate a person to engage in further harmful content online, but can also motivate offline violent extremism. It is also acknowledged, however, that the nature of these relationships between online and offline behaviour are complex, uncertain, and in need of further research.

To mitigate these harms, the Christchurch Principles offer a democratic model, which works from the assumption that democratic means and ends are the preferred strategy for reducing harm, because democratic institutions support the upholding of rights whose violation and non-realisation entail the most relevant forms of harm. As the American democratic theorist John Dewey (1987, p. 298) observed, "The fundamental principle of democracy is that the ends of freedom and individuality for all can be attained only by *means* that accord with those *ends*" [emphasis added].

In other words, the Christchurch Principles are concerned not only with upholding rights, but also upholding those democratic sentiments and institutions which create the best conditions for human rights to thrive. This creates responsibilities for states, businesses and civil society, all of which play various roles in the defence of, and active use of, democratic norms and practices.

The 2011 UN Guiding Principles on Business and Human Rights (UNGPs) provides an important precedent for the Christchurch Principles. In

particular, the latter proposes to adopt the former's *Protect, Respect and Remedy* (PRR) Framework. This sets out three pillars:

1. the duty of the State to *protect* against human rights abuses by third parties, including business enterprises, through appropriate policies, regulation and adjudication;
2. the corporate responsibility to *respect* human rights, which means that business enterprises should act with due diligence to avoid infringing the rights of others and to address adverse impacts with which they are involved; and
3. the need for greater access by victims to effective *remedy*, both judicial and non-judicial.

The UNGPs are relevant to the responsibilities of actors in the digital era, as recognised by the 2017 Global Network Initiative (GNI) Principles.¹ Yet the Christchurch Principles differs by extending the PRR Framework to democracy. That is, the Christchurch Principles assert that states, businesses and civil society have varying responsibilities to protect, respect and remedy those democratic institutional arrangements that create the most favourable conditions for human rights to be upheld. In this sense, the Christchurch Principles are concerned not only with rights, but also the underlying context of democratic norms and practices, which include trust, transparency, accountability, inclusivity, and non-deception.

In order to evaluate whether rights and democracy are being undermined, the Christchurch Principles uses the principle of equal participation as an evaluative standard.

This principle asserts that social, economic and political arrangements are ideally democratic when they support people to participate as equals in public life, whether in online or offline spheres. If a person's capability to contribute to public life is undermined, whether by online or offline causes, then this triggers the duties and responsibilities enshrined in the PRR Framework – that is, responsibilities to protect institutions that enable democratic participation, a duty of care to not undermine these institutions, and remedial interventions to address institutional vulnerabilities. The principle of equal participation has substantive affinities to Nancy Fraser's theory of participatory parity (Fraser 2000; *ibid.* 2005), and Amartya Sen's capability approach as applied to democratic participation (Sen, 2009).

In some cases, the threat to equal participation in democratic life can be defined by reference to canonical rights. For example, online incitements to violence against certain groups may constitute a credible threat to a person's "right to life, liberty and security" (UDHR, Article 3), and thereby imperil the person's willingness to participate in public life. This is a threat to democracy, but it can be addressed simply by reference to rights. For example, the 1995 Johannesburg Principles on National Security, Freedom of Expression and Access to Information

¹ See the GNI website <https://globalnetworkinitiative.org/gni-principles/>. The GNI Principles note: "The application of these Principles is informed by the UN Guiding Principles on Business and Human Rights ('UN Guiding Principles'), the 'Protect, Respect, and Remedy' Framework, and the OECD Guidelines for Multinational Enterprises."

declare that while expression of opinion and belief ought to be protected from “any sort of restraint, disadvantage or sanction” (Principle 5), this right is forfeited when such expression incites violence and therefore poses a threat to national security (Principle 6). Therefore the harms to democratic practice can be addressed through rights claims.

In other cases, however, the justification for PRR may not rest directly upon rights violations, nor threats to rights, but upon the degradation of norms and practices that are vital for the health of democratic institutions. For example, if a particular use of a digital technology is shown to undermine trust in democratic institutions, and this loss of trust impedes these institutions’ functioning, then the PRR Framework may be activated despite there being no “right to trust” to appeal to. However, it may be noted that these

degradations of democracy do, nevertheless, have an indirect relationship to certain rights, because it impedes the capacity for people to exercise, say, their right to partake in public affairs [UDHR, Article 21). In this sense, political rights and democratic institutions are mutually reinforcing.

These dynamics are encapsulated by Figure 1 below.

At the top, there is the PRR Framework, which allocates responsibilities for states, businesses and civil society. These organisations may form the basis for various kinds of society, but because the Christchurch Principles are a democratic model, they are overlaid by institutions that support human rights and democracy. As such, the substantive basis for a society to be democratic and rights-abiding is the extent to which the state, business and civil society uphold these cherished institutions.

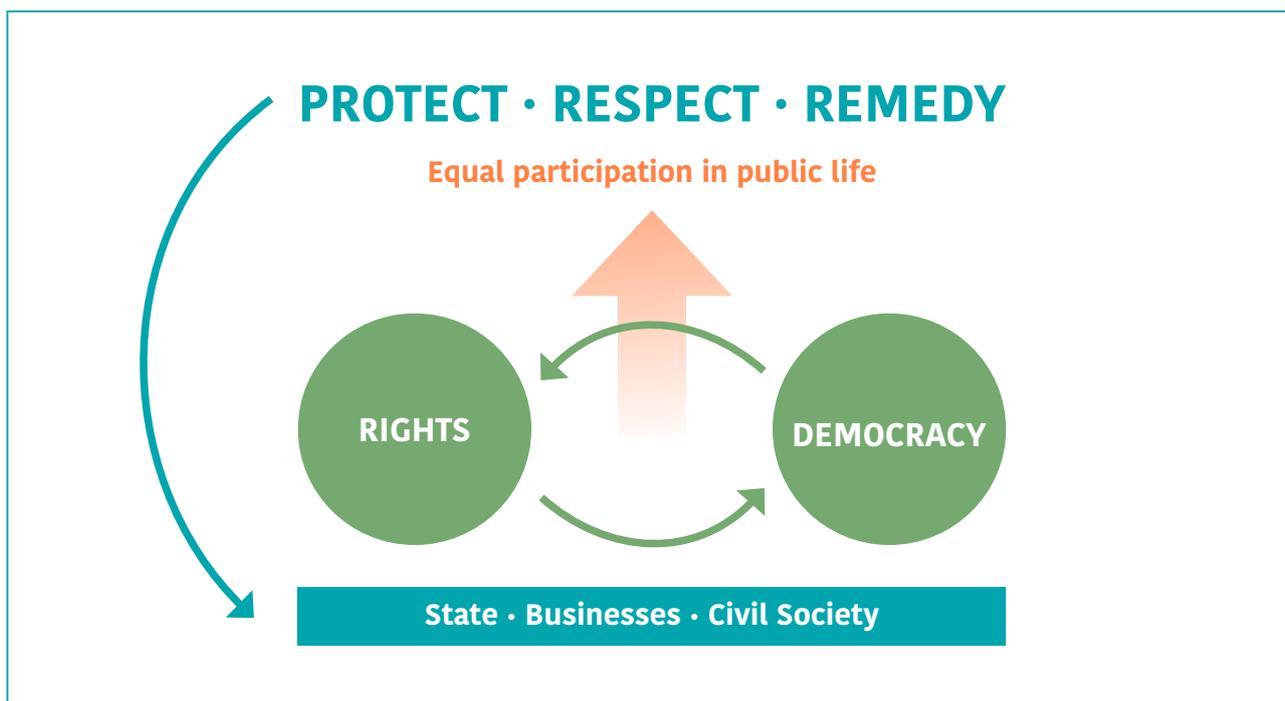


Figure 1

This model treats human rights and democracy as essentially interactive, by being mutually constraining and mutually reinforcing (Arat, 1991; Habermas, 1998; Waldron, 1999; Brettschneider, 2007; Christiano, 2008). In regards to mutual constraint, rights have long been conceived as a check on democracy, especially to prevent the tyranny of the majority, because rights establish inviolable rules that protect minority groups from majoritarian decision-making. On the other hand, democracy also serves as a check on liberal systems of rights, by constraining inequalities that accumulate under a system of strict non-interference. However, by mutually constraining each other's excesses, institutions of democracy and rights are mutually reinforcing, because the fulfilment of one creates favourable conditions for the fulfilment of the other.

When institutions of rights and democracy are functioning and well-balanced, then the ideal outcome is that all people within a political territory have the capability to participate as peers in public life (the principle of equal participation). However, if people do not have equal capabilities to participate in public life – for example, if certain groups of people are fearful of coordinated harassment, or if people lack access to online fora where debates of public importance occur – then responsibilities are activated under the PRR Framework. Furthermore, given the need to align democratic means and ends, it is noted that these responsibilities should be exercised in a manner that has

democratic legitimacy, or is itself a democratic exercise. For example, state action to protect rights and democratic institutions should enjoy consent through robust democratic endorsement, and businesses ought to exercise their respect for rights and democracy by integrating democratic norms and practices into their governance arrangements.

The Christchurch Principles have been informed by this framework, in order to get clearer about what is at stake from online hate, extremism and radicalisation. A rights-based perspective, which tends to be the sole focus of similar initiatives to date¹¹, is insufficient to capture the full breadth of impacts from the digital revolution. Digital technology companies typically appeal to rights, especially rights to freedom of expression, as a defence of business-as-usual. In the meantime, however, there is a growing literature which highlights the threats to democratic institutions posed by particular manifestations of digital technology (for example, Vaidhyanathan, 2018; Kaye, 2019; The Workshop, 2019; Couldry and Mejias, 2019; Digital, Culture, Media and Sport Committee, 2019; Zuboff, 2019). Thus, to properly account for the actual or potential harms of online hate and extremism, it is necessary to take this wider view, to explore not only direct threats to rights, but also indirect threats through the deterioration of democratic institutions which provide the most favourable conditions for human rights to flourish.

¹¹ The Manila Principles on Intermediary Liability (2015) <<https://www.manilaprinciples.org/>>; The Tshwane Principles on National Security and the Right to Information (2013), <<https://www.opensocietyfoundations.org/sites/default/files/tshwane-principles-15-points-09182013.pdf>>; The UN Guiding Principles on Business and Human Rights (2011) <<https://www.business-humanrights.org/en/un-guiding-principles>>; The Johannesburg Principles on National Security, Freedom of Expression and Access to Information (1996) <<https://www.article19.org/wp-content/uploads/2018/02/joburg-principles.pdf>>; Global Network Initiative Principles on Free Expression and Privacy (updated 2017) <<https://globalnetworkinitiative.org/gni-principles/>>.

Key concepts for the Christchurch Principles

The principle of equal participation

A well-functioning democracy is one where people within a political community can, and do, participate as equals in public life. It is a manifestation of the key democratic concept of political equality: ‘Political equality is achieved if all citizens have equal opportunity to participate in democratic processes and all voices are equally considered via procedures for achieving popular control of public decision-making’ (Birch, 2017, p.22; cf. Christiano 2008, pp. 75–130; Dahl 1989, pp. 109–115; Hyland 1995, pp. 51–75; Katz 1997, pp. 100–101; Saward 1998, pp. 49–67). The principle of equal participation serves as an evaluative standard for registering when reality is falling short of the democratic ideal.

There are various ways of articulating what contributes to an equal democratic standing. Nancy Fraser, in her theory of parity of participation, asserts that “justice requires social arrangements that permit all (adult) members of society to interact with one another as peers” (Fraser, 2003, p. 36). On her view, participatory parity is assured through certain conditions being met.

1. The objective condition relates to questions of economic inequality and dependency. It contends that equal participation is impeded by “deprivation, exploitation, and gross disparities in wealth, income and leisure time” (ibid., p. 36). This is the domain of struggles for *redistribution*.

2. The intersubjective condition requires that “institutionalized patterns of cultural value express equal respect for all participants and ensure equal opportunity for achieving social esteem”. In other words, people do not have equal standing if they are treated as “inferior, excluded, wholly other, or simply invisible” (Fraser, 2000, p. 113). This is the domain of struggles for *recognition*.

3. The political condition relates to “criteria of social belonging, and thus determining who counts as a member” of a political community (Fraser, 2005, p. 75). This includes voting rules, decision rights, and the boundaries of citizenship and political communities. This is the domain of struggles for *representation*.

In regards to the digital technologies, the question is to what extent these conditions are emboldened or inhibited through our newfound capacity to create and access online content. In certain ways, digital technologies are enabling; for instance, by creating opportunities for persons to participate in public debate online by overcoming earlier barriers of economic disadvantage or social prejudice. In other respects, online content may have not much bearing on certain conditions. For example, while the digital revolution most certainly has major implications for economic injustice (through tax evasion, employment practices, and wealth accumulation), the impact of online content on economic outcomes is likely to be subtle and indirect.

By contrast, the proliferation of hateful, aggressive and discriminatory content online may well have a prejudicial effect on people's respect and esteem, in ways that impede equal participation in public life. Similarly, coordinated campaigns of disinformation have interfered with democratic process to such a degree that political representation is under threat. The Russian targeting of African-Americans during the 2016 election with voter suppression techniques through social media (Shane and Frenkel, 2018) is an example of how political representation is harmed by online content.

Another way to conceive of the principle of equal standing is Amartya Sen's capability approach (Sen, 2009). Through this lens, the principle of equal standing may be conceived as *the capability to participate as an equal to others in public decisions*. Accordingly, this capability is a substantive freedom: political equality ensures that one can freely choose to participate on an equal standing with others. The foundation for such a capability is a set of *functionings* – that is, a set of pre-existing achievements which guarantee the opportunity to exercise certain freedoms. These functionings are sometimes phrased as 'beings' and 'doings'. For example, to be in good health, to be well-nourished, to be mobile, all have a bearing on whether one is capable of participating in public affairs.

To examine the principle of equal participation through the lens of the capability approach brings a level of concrete detail to its foundations. While Fraser focuses on the

normative dimensions of justice (recognition, redistribution and representation) that underlie political equality, Sen focuses on measurable achievements that may form the basis for an analysis of quality of life (Stiglitz, Sen and Fitoussi, 2009). While the idea of capabilities is also closely related to rights, it complements rights by being heedful of what people need in order to exercise their rights; for instance, a capability approach is attentive to the physical and mental heterogeneities among persons, variations in non-personal resources, environmental diversities, and different relative positions vis-à-vis others which mean that people's capability to exercise a right may differ, even when the right is protected and respected by key societal actors (Sen, 2005).

Digital communications technologies can enrich our capability to participate as equals in a variety of ways; for instance, through our ability to participate in discussions on public fora, to gain relevant information, to access diversity of opinion, and to provide marginalised groups with media to publish opinions and beliefs. However, it also challenges certain functionings. For example, our ability to be in public with dignity is compromised by certain dynamics of social media, especially the capacity for targeted harassment and shaming, which often targets certain already marginalised groups, such as women, people of colour and LGBTQ+ (two prominent examples are Gamergate and attacks on the all-female Ghostbusters film remake). This exposure to harassment, especially

for marginalised groups, creates the danger of discouraging certain groups from exercising their rights, from fear of abuse or harassment or targeted attack. This poses a threat to democracy by violating the principle of equal participation.

Canonical rights

UNGP's *Protect, Respect and Remedy* Framework is focused on human rights. It highlights the state duty to protect against human rights abuses committed by third parties, including business, through appropriate policies, regulation and adjudication. The corporate responsibility to respect human rights means acting with due diligence to avoid infringing on the rights of others, and addressing harms that do occur.

In regards to online content, there are direct and indirect implications for a variety of canonical rights, but the focus below is rights which online content has a direct connection to:

- **The right to freedom of expression (Article 19, UDHR; Article 19, ICCPR)**

- Digital communications technologies provide new means by which the right to freedom of expression can be exercised, especially for marginalised groups neglected by traditional media. However, design choices over internet platforms have implications, such as whether algorithms prejudice certain kinds of information, or lapses of moderation and privacy create unsafe online spaces by which people cannot exercise their right of expression.

- **The right to participate in public affairs (Article 21, UDHR; Article 25, ICCPR)**

- As digital technologies are increasingly integrated into ordinary life, there are greater potentials for impact, both positive and negative, on people's right to participate in the governance of their communities and societies. These include new opportunities for debate and dialogue, transparency and access to information, new forms of assembly and political organisation, and new vectors for discrimination and political interference.

- **The right to privacy (Article 12, UDHR; Article 17, ICCPR)**

- The internet enables new tools and mechanisms for states and private actors to monitor and collect information about individuals' communications and activities, which can violate the Internet users' right to privacy, or impede the free flow of information and ideas online by undermining people's confidence and security online.

- **The right to life, liberty and security (Article 3, UDHR)**

- Insofar as the internet is 'a driver and enabler for the process of radicalization' (New York Police Department, 2007, p.8) which can manifest in the offline world through acts of terrorism, violent extremism, discrimination and coordinated harassment, then the internet can have a causal connection to violations of the right to life, liberty and

security. This may have indirect impacts on democracy by discouraging people from exercising their rights to expression and political participation.

- **The right to equal dignity and non-discrimination (Articles 1 & 2, UDHR; Articles 2, 20 and 26, ICCPR)**

- The internet is a medium by which existing forms of discrimination and prejudice are manifested, from the offline to the online world. Digital communications technologies have also proven a novel instrument for exercising discrimination, especially through targeted campaigns of harassment and abuse. Moreover, digital technologies can themselves be a vector for discrimination, through choices in algorithm design which favour the interests of certain groups over others.

Non-canonical rights

- **The right to internet access**

- The right to internet access can be seen as a subsidiary of the right to freedom of expression, because the internet is an increasingly important medium through which this right can be exercised. A lack of access risks digital exclusion. Certain jurisdictions have already recognised a specific right to internet access, including Estonia, France and Finland.

- **The right to be forgotten**

- This refers to the deletion of online content about past events, such that people can “determine the development of their life in an autonomous way, without being perpetually or periodically stigmatized as a consequence of a specific action performed in the past” (Mantelero, 2013). The right to be forgotten has been operationalised in certain jurisdictions, such as the EU, South Korea and Argentina.

- **The right to sanctuary**

- The human need for inviolable refuge is captured in the right to asylum from persecution, but the digital era creates new needs for sanctuary. The greater interconnectivity of the digital world means that people are easier to track down, both in the offline and online world, for persecution or harassment. It also means we are subject to surveillance by both public and private organisations, the latter for commercial purposes. A right to sanctuary recognises the freedom to disconnect in order to secure one’s own autonomy.

- **The right to a future tense**

- This has been defined by Shoshana Zuboff as “the right to act free of the influence of illegitimate forces that operate outside of our awareness to influence, modify and condition our behaviour” (Zuboff, 2019, p. 195). It is a defence of individual autonomy

and sovereignty, especially in relation to the capacity for behavioural modification through online marketing and advertising, online nudges and suggestions, and the modification of one's choice architecture.

Democratic norms and practices

In their research, The Workshop adapted metrics used by the Economist Intelligence Unit for their Democratic Index report. The five key features of a healthy democracy considered in their research are:

- **Electoral process and pluralism:** Including whether elections are free, fair and trusted.
- **Active citizens:** Alert, informed citizens who are capable of making important moral judgements. This includes measures of equity, diversity and inclusion in representation.
- **Shared democratic culture:** Whether there's enough societal consensus, cohesion and willingness to compromise for a stable, functioning democracy.
- **Civil liberties and competitive economy:** A functioning competitive economy and civil society, including protection of human rights and a free, independent media.
- **Trust in authority:** Government, parliament, judiciary and public institutions are trustworthy and elected representatives are accountable to the people.

Some of the norms involved in support these practices, or which are involved in the democratisation of organisations, are:

- **Trustworthiness:** Subjective measurements of trust are useful, but do not tell us whether trust is justified or not. Hence the need to examine the richer concept of trustworthiness, which involves judgments of honesty, reliability and competency in the actor with whom we are placing or refusing our trust (O'Neill, 2017, *ibid.* 2018).
- **Transparency:** This requires public bodies to make certain types of information about their activities publicly available, either regularly or on demand, with some exceptions for non-disclosure and redaction. In regards to digital technologies, a greater level of transparency is desirable for algorithm design and the uses of data. Notably, transparency isn't sufficient, and needs to be supported by other norms; for instance, online contracts are often inaccessible to signatories due to their complexity and length, thus rendering transparency obscure.
- **Accountability:** This involves justifications for the use of power, combined with distributions of empowerments so that those subjected can sanction its use. As such, decisions which affect others can be held to account. Given the significant influence of digital technologies on people's private and public lives, there is an implicit requirement that these organisations are accountable for such activities.

- **Inclusivity:** This demands that procedures and structures are inclusive to a wide range of membership and/or representation, especially to include groups that are otherwise marginalised or misrepresented. This is important for digital technology companies given the narrow geographical and social context which informs the design of products, despite their global reach.
- **Reciprocity:** The social norm of responding to a positive action with another positive action, thus building cooperation and community.

Communicative action

Democratic participation requires “enlightened understanding... each citizen ought to have adequate and equal opportunities for discovering and validating (within the time permitted by the need for a decision) the choice on the matter to be decided that would best serve the citizen’s interest” (Dahl 1989, pp. 111–112). This is a demand that falls upon a society’s institutions to provide the intellectual resources – information, argumentation, values – for people to make well-informed decisions. This involves more than merely the rights to freedom of expression and access of information. After all, the unregulated provision of information also enables the flourishing of disinformation and misinformation; or speech acts that incite hatred, slander, defame, endanger or humiliate others; or information which is irrelevant, trivial or otherwise unhelpful. All this can degrade the quality of democratic decision making and participation.

As Onora O’Neill (2009, pp. 172–3) has argued: “Citizens need reliable information by which to judge what is done to them and in their name... Here living up to norms such as intelligibility, relevance, accuracy and honesty is likely to be important. We do not happily condone unintelligible or irrelevant communication by institutions; we do not pretend that it is mere self-expression... Nor do we condone inaccurate or dishonest institutional reports or financial accounts by classifying them as mere self-expression.”

Thus, information ought to satisfy criteria of successful communication action. This may involve:

- **Accessibility:** Information must “be capable of reaching its intended audience”, by being intelligible and relevant – i.e. it can be followed by its audience, and is relevant to the intended audience for the intended purposes (O’Neill, 2009, p. 175–6).
- **Assessability:** Information must be capable of being subjected to inquiry and critique in order to test its epistemic and ethical legitimacy, in order to protect against deception (O’Neill, 2009, p. 176–8). Transparency can contribute to assessability, by making “publicly available certain types of information about their activities, either regularly or on demand, apart from specific categories of reserved information” (ibid. p. 170), but transparency ought not be seen as sufficient without other norms (such as accessibility) being in play.

- **Accuracy:** Accuracy involves “the passion for getting it right” and “implies care, reliability, and so on, in discovering and coming to believe the truth” (Williams, p. 126–7).
- **Truthfulness:** Successful communication requires an intent to be truthful, which involves such values as sincerity and honesty. As O’Neill notes: “Truth-seeking demands discipline rather than mere self-expression, including disciplines such as experiment, careful inquiry, fact checking, and many others” (ibid. p. 172). This commitment to truthfulness is especially endangered by commercial motives such as profit-seeking, which may forego truthfulness in order to increase audience, which may occur through choices which do not favour truthfulness.

These norms have long been a part of media landscape, both through regulation and non-regulatory means. There are, for example, well established laws against perjury, defamation, calumny, vilification, slander, and libel. Similarly,

various jurisdictions have regulatory controls over false advertising. Non-government institutions have various rules in support of accuracy, such as peer-review process in academia, or codes of conduct in professional bodies. Media organisations have their own institutions for regulating content, both internal and external, such as codes of ethics, self-regulatory bodies such as press councils and ombudsman, and government regulations especially to control exceptions to freedom of expression (defamation, incitement of violence, etc.). The question for digital media communications is how such institutions might be developed in order to mitigate the harms caused to rights and democracy through such issues as online disinformation, misinformation, foreign government interference in democratic processes, the impact on public interest media, disengagement, attention hijacking, filter bubbles and echo chambers.

The Helen Clark Foundation released a report titled ‘**Anti-social media: reducing the spread of harmful content on social media networks**’ in May 2019. The Christchurch Principles build on the research and conclusions in this report. If you are interested to view the full report, please visit: <https://helenclark.foundation/social-media>

- Arat, Zehra F., 1991, *Democracy and Human Rights in Developing Countries*, Lynne Reiner Publishers.
- Birch, Sarah, 2017, 'Democratic norms, empirical realities and approaches to improving voter turnout', *RES PUBLICA*, 24(1), 9-30. <https://doi.org/10.1007/s11158-017-9380-3>.
- Brettschneider, Corey, 2007, *Democratic Rights: The Substance of Self-Government*, Princeton: Princeton University Press.
- Christiano, Thomas, 2008, *The Constitution of Equality*, Oxford: Oxford University Press.
- Couldry, Nick and Ulises Mejias, 2019, *The Costs of Connection: How Data is Colonizing Human Life and Appropriating it for Capitalism*, Stanford, CA: Stanford University Press.
- Dahl, Robert A., 1989. *Democracy and Its Critics*. New Haven, CT: Yale University Press.
- Dewey, John, 1987, 'Democracy is Radical', in (Jo Ann Boydston, ed.) *John Dewey: The Later Works, 1925-1953, Volume II: 1935-1937*, Carbondale: Southern Illinois University Press.
- Digital, Culture, Media and Sport Committee, 2019, *Disinformation and 'Fake News': Final Report*, Eighth Report of Session 2017-19, UK House of Commons.
- Ely, J. H., 1980, *Democracy and Distrust: A Theory of Judicial Review*, Cambridge, MA: Harvard University Press.
- Errington, Kathy and Claire Mason, 2019, *Anti-Social Media: Reducing the Spread of Harmful Content on Social Media Networks*, Auckland, <https://www.helenclark.foundation/wp-content/uploads/2019/05/thcf-social-media-report-min.pdf>
- Fraser, Nancy, 2000, 'Rethinking Recognition', *New Left Review* 3(3), pp. 107-118.
- Fraser, Nancy and Axel Honneth, 2003, *Redistribution Or Recognition?: A Political-philosophical Exchange*, London; New York: Verso.
- Fraser, Nancy, 2005, 'Reframing Justice in a Globalizing World', *New Left Review* 36, pp. 69-88.
- Habermas, Jürgen, 1998, 'Remarks on legitimation through human rights', *Philosophy and Social Criticism*, 24(2/3): 157-171.
- Hyland, James L., 1995. *Democratic Theory: The Philosophical Foundations*. Manchester: Manchester University Press.
- Katz, Richard S., 1997. *Democracy and Elections*. Oxford: Oxford University Press.
- Kaye, David, 2019. *Speech Police: The Global Struggle to Govern the Internet*, New York: Columbia Global Reports.
- Mantelero, Alessandro, 2013, 'The EU Proposal for a General Data Protection Regulation and the roots of the 'right to be forgotten'', *Computer Law & Security Review*, 29 (3), 229-235. doi:10.1016/j.clsr.2013.03.010.

- O'Neill, Onora, 2017, 'Intelligent Trust in a Digital World', *New Perspectives Quarterly*, 34(4): 27-31.
- O'Neill, Onora, 2018, 'Linking Trust to Trustworthiness', *International Journal of Philosophical Studies*, 26:2, 293-300, DOI: 10.1080/09672559.2018.1454637
- Saward, Michael. 1998. *The Terms of Democracy*. Oxford: Polity.
- Sen, Amartya, 2005, 'Human Rights and Capabilities', *Journal of Human Development*, 6(2): 151-166.
- Sen, Amartya, 2009, *The Idea of Justice*, London & New York: Penguin.
- Shane, Scott and Sheera Frenkel, 2018, 'Russian 2016 Influence Operation Targeted African-Americans on Social Media', *New York Times*, 17 December 2018, <https://www.nytimes.com/2018/12/17/us/politics/russia-2016-influence-campaign.html>
- Stiglitz, Joseph; Sen, Amartya and Fitoussi, Jean-Paul, 2009, *Report By The Commission On The Measurement Of Economic Performance And Social Progress*. Paris: Commission on the Measurement of Economic Performance and Social Progress.
- United Nations, 2011, *United Nations Guiding Principles on Business and Human Rights, Implementing the United Nations "Protect, Respect and Remedy" Framework*, New York and Geneva: United Nations.
- Vaidhyanathan, Siva, 2018, *Anti-Social Media: How Facebook Disconnects Us and Undermines Democracy*, Oxford: Oxford University Press.
- Waldron, Jeremy, 1999, *Law and Disagreement*, Oxford: Oxford University Press.
- The Workshop, 2019, *Digital Threats to Democracy*, Wellington: The Workshop.
- Williams, Bernard, 2002, *Truth and Truthfulness: An Essay in Genealogy*, Princeton, NJ: Princeton University Press.
- Zuboff, Shoshana, 2019, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, London: Profile Books.

THE
**Helen
Clark**
FOUNDATION

