Uneven Platforms: The Press, Social Media, Search Engines and Freedom of Expression

PAUL TAYLOR*

Abstract

Digital platforms have emerged as giants in the diverse and rapidly developing communication infrastructure, which gives access to news and other information, as well as enabling individuals and businesses to communicate privately, within communities or to a universal audience. It is argued that digital platforms differ from conventional publishers, as they have little control in the content they carry and present to users. Content is posted on a scale which is unmanageable, so digital platforms rely on internal policies and a cortege of automated and human processes to filter harmful content. They also capitalise on user data to exploit market opportunities. The question now facing digital platforms, particularly as they suffer reputationally from allegations of anti-competitive conduct, violation of user privacy, and political bias (among other distortions), is whether the standards they apply when curating content and unsubscribing users are compliant with human rights law. This article argues that they are not, and that the appropriate standards are those set under the International Covenant on Civil and Political Rights ('ICCPR'). Contracting States to the ICCPR should take responsibility for ensuring those standards are effectually secured. In light of the capacity of digital platforms in Australia to wield market power and technological prowess in the promotion of their own commercial interests, they are considered inappropriate candidates for the responsibilities they currently bear in content moderation. The lack of clear adherence to ICCPR standards in Australia hampers the call for their application in this context made by the UN Special Rapporteur on freedom of expression, and others.

I INTRODUCTION

'O, it is excellent To have a giant's strength; but it is tyrannous To use it like a giant.'[1]

The prodigious growth in the scale and power of digital platforms, such as Facebook and Google, is not simply a function of consumer demand for

*Senior Research Fellow, TC Beirne School of Law, University of Queensland; Visiting Fellow, Centre for International and Public Law, College of Law, Australian National University; Adjunct Professor, School of Law, The University of Notre Dame Australia. The author acknowledges the generosity of the Estate of Douglas Slatter and Elizabeth Chambers in the provision of funding which supported the writing of this article.

[1] William Shakespeare, *Measure for Measure*, Act 2 Scene 2, 130 (Isabella).

their services; it is also the result of their expansion into neighbouring markets, for example, by commercially exploiting the data they hold on users.[2] Social media platforms have become indispensable tools for maintaining private and business contacts in enclosed or open fora on a scale that was previously unimaginable. Public dependence on the famous tech companies of today and the popularity of their offerings quickened their market ascendancy. Yet, public confidence in particular tech companies has been dented by large-scale anti-trust investigations, the disclosure of serious misuses of data in violation of user privacy, and accusations of political bias and voter manipulation, among other reputational setbacks. Google, in particular, has demonstrated both its market dominance and abuse of dominant position in its core services by leveraging power in other markets.[3]

All this comes at a time when more is entrusted to tech companies. The responsible operation of a social media platform demands appropriate curation of content. The task of regulating content has fallen to the platforms themselves by default, and they have responded by developing policies which bind users to 'community standards' and similar norms. This prompts the questions of whether this form of private regulation of content adheres to human rights standards, and whether the platforms themselves abide by democratic principles which underpin human rights protection. It also spurs an assessment of whether governments are sufficiently interventionist in their supervision and regulation of the gatekeeping responsibilities of such platforms. The priorities of these platforms include maximising profitability and shareholder yields, attracting and consolidating customer loyalty, and developing opportunities in available markets, sometimes, it appears, even at the expense of the rights of their own customers.

---

[2] For the potential to monetise other digital data left by users: see Engin Isin and Evelyn Ruppert, *Being Digital Citizens* (Rowman & Littlefield, 2015) 103–5. They describe such digital traces as 'the new oil to be tapped'.

[3] Google fined €2.42 billion for abusing dominance as search engine by giving illegal advantage to own comparison-shopping service: *Summary of Commission decision of 27 June 2017 relating to a proceeding under Article 103 of the Treaty on the Functioning of the European Union and Article 54 of the EEA Agreement (Case AT.39740 – Google Search (Shopping))* [2018] OJ C 9/11. Google fined €4.34 billion for illegal practices regarding Android mobile devices to strengthen dominance of Google's search engine: *Summary of Commission Decision of 18 July 2018 relating to a proceeding under Article 102 of the Treaty on the Function of the European Union and Article 54 of the EEA Agreement (Case AT.40099 – Google Android)* [2019] OJ C 402/08; For an analysis of characteristics of information (as a good) and digital platforms in the application of competition law: see Beata Mäihäniemi, *Competition Law and Big Data: Imposing Access to Information in Digital Markets* (Edward Elgar Publishing, 2020).

The enquiry undertaken in this article falls within a narrow compass. It is not about the regulation of the Internet as a whole,[4] or of particular emerging technologies.[5] It is confined primarily to the legal regulation of digital platforms as they curate content, and surface news and other content for users. Morgan and Yeung depict a narrow concept of 'regulation', corresponding with the one adopted here, centring on 'deliberate attempts by the state to influence socially valuable behaviours which may have adverse side-effects, by establishing, monitoring and enforcing legal rules'.[6] It contrasts with their broad concept of regulation (which could be regarded as the target of the legal regulation of digital platforms under discussion), encompassing social control, following:

> the emergence of non-state institutions, including commercial enterprise and non-governmental organisations, that operate both as a source of social influence and a forum in which public deliberation may occur.[7]

The present focus is the State regulation of the operations of digital platforms, as they undertake their content curation and offer other services, to ensure that it does not unjustifiably impact the freedom of expression of users. It is not based on any notion that such non-State entities are directly bound by human rights obligations (which would be controversial), but instead assumes that it is the State's responsibility to secure human rights, even against private sources of violation. It is not necessary to go as far as Andrew Clapham did, when he argued that 'the existing general rules of international human rights law, created and acknowledged by states, now fix on non-state actors so that they may be held accountable for violations of this law'.[8] He candidly acknowledged that those who insist that States are the only bearers of human rights obligations are likely to dismiss his argument as incorrect.[9] The more conventional approach, focusing on State responsibility, is sufficient for the purposes of this article. Of importance

---

[4] See eg Lilian Edwards, *Law, Policy and the Internet* (Hart Publishing, 2018) for discussion of e-commerce (including the role and responsibilities of online intermediaries) and privacy, data protection and online crime in the UK context.

[5] For coverage of the regulation of a broad range of technologies (including biotechnologies, nanotechnologies and neurotechnologies), with specific focus on different policy spheres see Roger Brownsword, Eloise Scotford and Karen Yeung (eds), *The Oxford Handbook of Law, Regulation and Technology* (Oxford University Press, 2017). Karen Yeung was also the Rapporteur for the Council of Europe Study on human rights dimensions of automated data processing and different forms of artificial intelligence: Karen Yeung, Special Rapporteur, *A study of the implications of advanced digital technologies (including AI systems) for the concept of responsibility within a human rights framework*, DGI(2019)05 (2019).

[6] Bronwen Morgan and Karen Yeung, *An Introduction to Law and Regulation Text and Materials – The Law in Context Series* (Cambridge University Press, 2007) 3.

[7] Ibid 4.

[8] Andrew Clapham, *Obligations of Non-State Actors* (Oxford University Press, 2006) 29. For discussion of the doctrine of state responsibility for regulating private actors to observe human rights: see also Danwood Chirwa, 'The Doctrine of State Responsibility as a Potential Means of Holding Private Actors Accountable for Human Rights' (2004) 5 *Melbourne Journal of International Law* 1.

[9] Clapham (n 8) 29.

to the present discussion is art 2(1) of the *International Covenant on Civil and Political rights* ('*ICCPR*').[10] It obliges States Parties not only to respect the rights enshrined in the *ICCPR*, but to also ensure them, to everyone within their jurisdiction, including where the threat to those rights is a private source. Freedom of expression is to be guaranteed under art 19 of the *ICCPR*. When commercial entities with vested interests unduly curtail individual expression, promote news asymmetrically, or limit access to information, State intervention becomes appropriate as an art 19 matter.

The power digital platforms wield is of concern not only as a matter of human rights law, but also competition law. This article begins by considering briefly (in section II) the importance of the 'free press' as a key historic contribution to the conditions of democracy which enable the enjoyment of individual freedoms. Section III considers the implications of the economic and technological capabilities of a handful of players which operate online search engines, social media platforms, news media services and provide online display advertising. The purpose is to demonstrate that such power carries with it the capability to introduce distortions in public access to information critical to opinion formation, and the ability to apply self-selected rules of censorship without traditional forms of accountability. Corrective measures are suggested in section IV, to bind platform operators to the objective and universal standards established by art 19 of the *ICCPR*. Contracting States to the *ICCPR*, including Australia, are best placed to secure this, and are already charged with that task as a matter of international obligation. There is, however, no substantive protection in Australia matching that required by art 19. This art draws on recent observations of the UN *Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*,[11] to suggest that it is nevertheless unsustainable for Australia to acquiesce in its responsibility by devolving it to digital platform owners in a form of private self-regulation when they are driven by conflicting priorities.

II A 'FREE PRESS'? ALL MEDIA IN THE SERVICE OF DEMOCRATIC STANDARDS

A free and independent press is fundamental to the efficacy of modern democracy, civic participation, enhanced social welfare, and even higher life satisfaction.[12] For many countries, press freedom came at great cost. In

[10] *International Covenant on Civil and Political Rights*, opened for signature, 16 December 1966, 999 UNTS 171 (entered into force 23 March 1976) ('*ICCPR*').

[11] The Special Rapporteur's mandate was created in 1993, inter alia, to examine issues affecting freedom of expression, to report and to make recommendations to better promote and protect the freedom.

[12] Christopher Ambrey et al, 'On the Confluence of Freedom of the Press, Control of Corruption and Societal Welfare' (2016) 128(2) *Social Indicators Research* 859–80.

England, it was secured principally by the removal in 1695 of legislative licensing controls on the press, following centuries of stringent copyright and censorship measures in different forms. Pressure for this had been mounting under the influence of John Milton's Civil War pamphlet, *Areopagitica*, although John Locke's campaigning efforts in the 1690s were conclusive. At the same time, the monopoly which the Stationers' Company held over printing was ended. In the years that followed, the history of democracy saw the electoral system develop along discernible party-political lines, heralding a new era of public engagement in politics. Parliamentary opposition began to take shape and journalism assumed a decidedly partisan character. However, MPs at that time perceived potential harms in a free press, compared with speech in Parliament which could be corrected in real-time if it was false or dangerous.[13] The experience of some countries shows that a free press once won may easily be lost. In Sweden, press freedom was constitutionally enshrined in 1766, alongside the principle of public access to official records. This was the initiative of Anders Chydenius, an outspoken advocate of natural equality, among other human rights. He wrote that '[n]o evidence should be needed that a certain freedom of writing and printing is the backbone of a free organisation of the state.'[14] Yet by 1774, following a coup d'état, the freedom was eviscerated by the King's arrogation of the power to determine what may be printed or officially disclosed, with non-observance punishable in extreme cases by execution. Press freedom was only restored in Sweden's Constitution 35 years later.[15] The US Constitution's First Amendment gave formal acknowledgement to press freedom, in conjunction with freedom of speech, as 'one of the greatest bulwarks of liberty'.[16] Countries inevitably differ in their constitutional formulation of these freedoms. In Australia press freedom and freedom of expression are not constitutionally secured. Prominent in the protection available in Australia is the implied freedom of political communication, which is 'limited to what is necessary for the effective operation of that system of representative and responsible government provided for by the *Constitution*'.[17] The implied freedom is supported by the common law freedom that exists generally in the absence of contrary regulation.[18] Together this may be described as 'free speech', but it does not correspond with the model of art 19 of the *ICCPR* which treats freedom of expression

[13] For an historical overview see Tom O'Malley, 'Regulation of the Press,' in Martin Conboy and John Steel (eds), *The Routledge Companion to British Media History* (Routledge, 2014) 228–39.
[14] Anders Chydenius, *Memorandum Concerning the Law for Freedom of Information* (1776).
[15] For a detailed historical account see Bertil Wennberg and Kristina Örtenhed (eds), *Press Freedom 250 Years: Freedom of the Press and Public Access to Official Documents in Sweden and Finland – a Living Heritage from 1766* (Sveriges Riksdag, 2018).
[16] James Madison, *Proposed Amendments to the Constitution*, Annals of Congress, House of Representatives, 1st Congress, 1st Session, June 8, 1789 (Gales and Seaton, 1834) 451–53.
[17] *Lange v Australian Broadcasting Corporation* (1997) 189 CLR 520, 561.
[18] '[E]verybody is free to do anything, subject only to the provisions of the law': Ibid 564.

as a personal right, of significant breadth, and permits restraint on the freedom only on specific grounds, and only when particular disciplines are observed. art 19 applies its protective standards against restriction whether the source is public or private, and whether it is legislative in origin or not. The demands of art 19 are discussed further below. Indeed, 'protection of free speech at the Commonwealth level essentially dates back to 1992 and is very limited compared with the equivalent protection under international law'.[19]

At international law press freedom is intended to operate in combination with the full suite of human rights to be guaranteed to the human person by any modern democratic State. It exists as a key component of every person's freedom of expression in art 19 to 'seek, receive and impart information and ideas of all kinds'. Although journalists do not possess special powers of investigation, they do enjoy certain privileges (notably 'source privilege'), and restrictions on journalistic expression are regarded with particular sensitivity.[20] Today the public still depends on, and assumes the existence of, press freedom in its broader consumption of mass electronic media. However, for that freedom to be effective certain preconditions must be satisfied. These might be categorised as: structural, functional and ethical. The structural dimension requires that a reasonable diversity of political opinion is broadly reflected in a representative plurality of media content. The functional aspect concerns the operation of the press in bringing significant matters to public attention and, as watchdog, holding individuals and institutions who misuse their power to account, especially so that significant failings by those in public office will not go unobserved. The ethical element requires that news content be fair and accurate, and not distorted.

Journalists are conveyors of information that is beyond the normal reach of the general public. Much of this information, especially news, is presented through the medium of critical professional analysis. Public trust in journalism need not be naive or unqualified. It is properly informed by an awareness that content is influenced by the ownership of the news source (in Australia dominated by private corporations, such as News Corporation and Fairfax Media on one hand, the publicly funded ABC on the other), each appealing to different politically aligned audiences and staffed

[19] Monash University Castan Centre for Human Rights, Submission No 18 to Australian Law Reform Commission ('ALRC'), *Traditional Rights and Freedoms—Encroachments by Commonwealth Laws* (Report No 129, March 2016), cited in ALRC, *Traditional Rights and Freedoms—Encroachments by Commonwealth Laws* (Report No 129, March 2016) ch 4 [4.35].

[20] In Australia, this privilege is enshrined in the *Evidence Act 2008* (Cth) s126K (journalist privilege relating to identity of informant). Among UN sources, the Human Rights Committee's *General Comment No 34* acknowledges as a component of the freedom of expression that journalists are entitled not to disclose their sources: *General Comment No 34, Article 19, Freedoms of opinion and expression*, 19th sess, UN Doc CCPR/C/GC/34 (12 September 2011) [45] ('GC 34').

according to organisational preferences for those supporting the brand. In this context, all media organisations select which stories will be covered, who will be interviewed, and which commentators will be chosen for comment. Of itself this should not be regarded as 'bias'. For present purposes, that term is confined to bias which offends journalistic standards with some prevalence. Even where bias does exist, its harms are generally mitigated where political or other weighting is self-evident to the educated consumer, who is able to receive it more or less critically.[21]

Some commentators are particularly sceptical about the media. Herman and Chomsky, for example, consider the media's role in democracy to be mythical, even undemocratic. They suggested that all media output is system-supportive propaganda, reflecting the interests of those wielding power. They depicted the reliance of mass communication media on market forces, internalised assumptions, and self-censorship (rather than overt coercion), in which various editorially distorting filters are applied to news reporting.[22] As described by Kemp, their 'propaganda model'

> essentially claims that media output in free-market-based media systems is propaganda because it heavily reflects the interests of those already in power … While control of the system is not overt or coordinated, markets have inherent incentives and disincentives that promote certain types of output and punish others. As such, certain agendas and frames dominate while others are marginalised. The authors argue that ideas around objectivity, representation and accountability, which are central to the media's alleged democratic role, are largely myths. In fact they argue that the media system works against democracy and is actually undemocratic.[23]

This model explains how money and power have the capacity to select favourable news content, marginalise dissent, and enable both government and dominant private interests to do their public messaging. More recently, the scepticism expressed by Lebovic was framed around how the right of free speech in America failed to guarantee press freedom, in departure from an ideal of free public access to accurate information, including in the news.[24]

---

[21] See Richard Paul and Linda Elder, *How to Detect Media Bias & Propaganda* (Foundation for Critical Thinking, 2004) 2: 'Democracy can be an effective form of government only to the extent that the public (that rules it in theory) is well-informed about national and international events and can think independently and critically about those events. If the vast majority of citizens do not recognize bias in their nation's news; if they cannot detect ideology, slant, and spin, if they cannot recognize propaganda when exposed to it, they cannot reasonably determine what media messages have to be supplemented, counter-balanced, or thrown out entirely'.

[22] Edward Herman and Noam Chomsky, *Manufacturing Consent: The Political Economy of the Mass Media* (Pantheon Books, 1988) 306.

[23] Geoff Kemp, *Politics and the Media* (Auckland University Press, 2016).

[24] Sam Lebovic, *Free Speech and Unfree News: The Paradox of Press Freedom in America* (Harvard University Press, 2016).

The advent of search engines, social media and of new platforms for news dissemination demands a fresh assessment of the bases of power and the actors wielding influence to determine the extent to which digital platform operators involved have the capacity and motivation to serve or to distort public messaging. There are three types of power available to them, which will be considered in the next section. They are separate but interrelated: market power in the competition law or anti-trust sense; technocratic power, often hidden in its impact, for example, in news feed selection and ranking of information sources; and the power of curation, to make and enforce rules of censorship without adequate accountability. The concept of a free press traditionally implies freedom from government interference. The challenge today is that digital platforms represent a private source of interference in the way in which they exert or have capacity for such influence contrary to human rights standards.

III *QUIS CUSTODIET IPSOS CUSTODES*?

When reporting to the UN Human Rights Council, the *Independent Expert on the promotion of a democratic and equitable international order*, Alfred-Maurice de Zayas, observed that the above question, 'who will watch the watchmen?', posed in the *Satires* of Juvenal almost 2000 years ago,[25] 'remains a central concern of democracy, since the people must always watch over the constitutional behaviour of the leaders and impeach them if they act in contravention of their duties'.[26] The principle should not be confined to democratic leaders, but, it is suggested, also to those who wield any of these sources of power (market power, technocratic power and the power of curation) with adverse consequences for democratic freedoms.

In Australia there is an extremely vigilant watchman over market distortions in the form of the competition regulator, the Australian Competition and Consumer Commission ('ACCC'). It has a clear mandate and effective machinery to address anti-competitive conduct, to answer structural issues presented by mergers, and to support the regulatory framework for particular industry sectors. The ACCC recently completed a report, as part of its Digital Platforms Inquiry, into the impact on competition in media and advertising services markets of digital search engines, social media platforms, and digital content aggregators.[27] The observations made in the report provide a number of insights into the dynamics of that sector which speak to the economic and technological power of digital platforms and their commercial priorities. The human

[25] Decimus Junius Juvenalis, *Satire* VI, lines 347–48.
[26] Human Rights Council, *Report of the Independent Expert on the promotion of a democratic and equitable international order*, UN Doc A/HRC/24/38 (1 July 2013) [52].
[27] Australian Competition and Consumer Commission, *Digital Platforms Inquiry* (Final Report, 26 July2019) ('ACCC Final Report').

rights implications for free speech include the suitability of such private entities as gatekeepers of digital content when setting free speech standards for users and curating content. It is argued here that they do not have the expertise, impartiality, or capacity for that function, and, in any event, the authority and responsibility to uphold human rights standards for all within the jurisdiction belongs to and should remain with the State. If Juvenal's anxiety, as popularly understood, applies to misuse of authority by the State, how much greater should our concern be when power is left in the hands of companies operating in their own interests in the private sector?

A *Competition law and human rights law*

Competition law and human rights law intersect at various points. When competition law investigations and proceedings first became commonplace, they produced fair trial concerns for those accused of competition law infringement.[28] More recently, human rights law has combined with competition law and broader public welfare concerns, for example, where access to essential medicines seemed to be impeded by obstacles to competition posed by pharmaceutical companies.[29] The ownership of mass media has long been understood as belonging to the province of merger control, focusing on the structural loss of competition that may ensue from excessive concentration, or the capacity that is enhanced for misusing dominant market power. The ownership of media outlets also provokes human rights concerns for freedom of expression, even in the private sector where specific government control is not exerted over media content. The priority of the UN Human Rights Committee, which monitors compliance with the *ICCPR*, is to foster media plurality, and to avert the risk that diversity of sources and views might be adversely impacted by monopolies and other concentrations.[30] Behavioural issues

---

[28] See eg, Arianna Andreangeli, *Competition Law and Human Rights: Striking a Balance Between Business Freedom and Regulatory Intervention* (Edinburgh Research Explorer, 2009) 6–8.

[29] Kwanghyuk Yoo, 'Interaction of Human Rights Law and Competition Law: The Right to Access to Medicines and Consumer Welfare in the Pharmaceutical Sector' (2018) 43 *Vermont Law Review* 123. Note, however, that there is no right to health under the *ICCPR*, even if other instruments support one.

[30] GC 34 (n 20) [40]. For Human Rights Committee concern about concentration of media ownership and influence by political and private interests that may not reflect public interest see eg its observations in the following reports: Human Rights Committee, *Concluding observations on the fifth periodic report of Argentina*, 117th sess, UN Doc CCPR/C/ARG/CO/5 (10 August 2016) [35], [36]; Human Rights Committee, *Concluding observations on the fifth periodic report of Moldova,* 118th sess, UN Doc CCPR/C/MDA/CO/3 (18 November 2016) [31] ('*Moldova report*'); Human Rights Committee, *Concluding observations on the fourth periodic report of Bulgaria*, 124th sess, UN Doc CCPR/C/BGR/CO/4 (1 November 2018) [37] ('*Bulgaria report*'). For allocation of public funding for the media and journalists in a transparent and non-discriminatory manner see eg *Bulgaria report* [38]. For recommendations that efforts be made to increase media pluralism and the diversity of views and information accessible to the public see eg *Moldova report* [32]; *Bulgaria report* [38]. For other aspects of media independence see e.g. Human Rights Committee, *Concluding observations on the fourth periodic report of Azerbaijan,*

which impact competition are increasingly of concern to competition regulators in the dynamic and rapidly changing environment in which digital platforms operate. The challenge, as Ioannis Lianos recently observed, is that competition law in the digital era needs to take account of new processes of value generation and capture in the era of digital capitalism, rather than rely on established competition law frameworks.[31] Rising to that challenge in Australia, the ACCC highlighted a variety of issues emerging from the market power and practices of those providing social network services in their core and neighbouring markets. This generates concern (which is beyond the ACCC's remit) for the suitability of such entities to be entrusted with content curation where it could significantly impact freedom of expression. One aspect of that is regulatory compliance track record of those companies. In other jurisdictions Microsoft, Twitter, and Facebook have attracted competition scrutiny over recent decades.[32] In late 2019, social media platforms were the subject of fresh anti-trust investigation in the US, including whether Facebook stifled competition.[33]

B *The moderation of digital content by the key market players in Australia*

As the ACCC's report explains, Google and Facebook are the two largest digital platforms in Australia, and unavoidable partners for Australian

---

118th sess, UN Doc CCPR/C/AZE/CO/4 (16 November 2016) [36] (arbitrary interference with media freedom, including revocation of broadcast licences allegedly on political grounds; politically motivated proceedings against independent media outlets; and alleged financial pressure on the independent newspaper); Human Rights Committee, *Concluding observations on the third periodic report of Bosnia & Herzegovina*, 119th sess, UN Doc CCPR/C/BIH/CO/3 (13 April 2017) [37] (the media was subjected to excessive influence from governments, political parties and private interest groups); Human Rights Committee, *Concluding observations on the third periodic report of Serbia*, 119th sess, UN Doc CCPR/C/SRB/CO/3 (10 April 2017) [38] (ongoing public influence exercised on some media); Human Rights Committee, *Concluding observations on the second periodic report of Turkmenistan*, 119th sess, UN Doc CCPR/C/TKM/CO/2 (20 April 2017) [42] (absence of a genuine independent media despite legislation); Human Rights Committee, *Concluding observations on the fifth periodic report of Belarus*, 124th sess, UN Doc CCPR/C/BLR/CO/5 (1 November 2018) [49] (executive power to shut down media outlets and use of warnings to media outlets that had a chilling effect); Human Rights Committee, *Concluding observations on the sixth periodic report of Hungary*, 122nd sess, UN Doc CCPR/C/HUN/CO/6 (9 May 2018) [57] (the regulatory bodies lacked sufficient independence to perform their functions).

[31] Ioannis Lianos, 'Competition Law for the Digital Era: A Complex Systems' Perspective' (Research Paper No 6/2019, Centre for Law, Economics and Society, University College London, August 2019) 8.

[32] Spencer Weber Waller, 'Antitrust and Social Networking' (2011) 90(5) *North Carolina Law Review* 1771.

[33] Kari Paul, 'Facebook and Google antitrust investigations: all you need to know' *The Guardian* (online, 7 September 2019) <https://www.theguardian.com/technology/2019/sep/06/facebook-google-antitrust-investigations-explained>.

news media businesses.[34] They are therefore of special importance given their potential effects on media plurality. They operate multi-sided platforms, among other things, generating advertising revenue from data sets about their users, and by providing services to consumers 'freely' in exchange for consumers' data, which they monetise by exploiting it in other markets. There they offer highly personalised and accurately targeted advertising profiles to advertisers. In Australia, Google has substantial market power in the supply of general search services and search advertising services. Facebook has substantial market power in the supply of social media services and display advertising services. Each platform has substantial bargaining power in its dealings with news media businesses. In the case of Facebook, this derives from providing media businesses with a key distribution channel to target particular demographic groups, in return for extending the Facebook offering with news content. The ACCC was astute to the potential for Google or Facebook to manipulate their algorithms or alter the display of content on their search engine results page or newsfeed to affect traffic to websites.[35]

Digital platforms tend to amplify the opportunity for information disorder by those who use them, in the form of misleading headlines, doctored photographs, misleading news commentary, and factual mistakes in the media. The ACCC found it difficult to determine how many incidents of unreliable news were egregious or serious; most adult Australian news consumers witnessed issues they considered serious, but only a minority complained.[36] On the other hand, consumers may consider news content that they disagree with to be unreliable or of poor quality, and report it as information disorder.[37] To put the scale of information disorder in context, the ACCC found evidence suggesting that exposure to disinformation and misinformation may be confined to 'heavy social media users who dig deeper into the long tail of news outlets beyond the mainstream'.[38] It observed from recent studies that exposure to false news was driven by consumers' demand for a variety of news; that real news audiences dwarfed false news audiences; and that individual users who accessed false news sites spent around half as much time per visit on false news sites relative to real news sites. This suggests that the audiences for false news arrive at sources for this content through a desire for more variation in media sources, and that social media facilitates this.[39]

The ACCC found that digital platforms have considerable influence in shaping the news viewed by Australian consumers and, importantly, perform curatorial functions when surfacing information. It noted that the

[34] ACCC Final Report (n 27) 4.
[35] Ibid 529–530.
[36] Ibid 354.
[37] Ibid.
[38] Ibid 356.
[39] Ibid 356–7.

atomisation of media content and the risk of misinformation and disinformation being spread on digital platforms hampered consumers' ability to evaluate the veracity, trustworthiness, and quality of the news content they receive online. The personalisation of content to users by digital platforms can also obscure the exact level of disinformation or mal-information presented to them. The ACCC rightly observes that '[w]hile public interest journalism contributes to a healthy democracy, disinformation and mal-information does the opposite'.[40]

At present, a great deal of public confidence is placed in digital platforms concerning trustworthy news sources. 'Any bias or preference given effect by the platform, either human or algorithmic, could influence the information presented to consumers.'[41] The ACCC observed that the delivery of news through messaging relies entirely on hosting services by private digital platforms like Facebook.

> This messaging format is at risk of interference from the hosting service, which could influence, rank, moderate or charge for its use by news media businesses. If the delivery of news through private messaging becomes more popular over time, the content, accountability and control of these services should be considered as part of any broader updates to the regulatory frameworks governing media and journalism.[42]

Artificial intelligence ('AI') is capable of further enabling digital platforms to filter for disinformation, misinformation and mal-information. AI processes can evaluate how well news stories match their headlines and can be used to de-prioritise content identified as spreading low-quality news. Machine learning is already used by Google News to prioritise high-quality news sources, and by Facebook to identify and block fake accounts, and those who spread spam and fraudulent material. The sensitivity in this context is the prospect that 'AI systems that exhibit statistical biases in their models or algorithms can result in actions that cause undesirable, unequal and/or unfair outcomes', as a study by the Centre for Media Transition

---

[40] Ibid 358. On the proposition that the fake news audience is small and less influential than might be imagined see Jacob Nelson and Harsh Taneja, 'The Small, Disloyal Fake News Audience: The Role of Audience Availability in Fake News Consumption' (2018) 20(10) *New Media & Society* 3720.

[41] ACCC Final Report (n 27) 364. The contemporary dependence on algorithmic systems poses threats as well as benefits which are evaluated in detail, including in the context of existing public sector applications, and with regulatory considerations, in a wide range of contributions in Karen Yeung and Martin Lodge (eds), *Algorithmic Regulation* (Oxford University Press, 2019). For studies based on the experiences of human moderators which disclose the need to achieve an appropriate balance between human and machine decision-making see: Shagun Jhaver et al, 'Human-Machine Collaboration for Content Regulation: The Case of Reddit Automoderator' (2019) 26(5) *ACM Transactions on Computer-Human Interaction* Article 31; Minna Ruckenstein, Linda Lisa and Maria Turunen, 'Re-humanizing the platform: Content moderators and the logic of care' (2019) 22(6) *New Media & Society* 1026.

[42] ACCC Final Report (n 27) 522.

pointed out.[43] This may simply be produced by the unconscious bias of the programmers of AI software. Chatbots (applications that enable on-line chats) also pose threats of a democratic nature, when interactions with them create artificial consensus.

The purpose in airing these aspects of the ACCC's report is to make the point that digital platforms have extraordinary capability, whether motivated nobly or otherwise, to tilt imperceptibly what the user 'eyeballs', to apply unseen human and algorithmic processes to promote or downgrade newsfeeds and other content, and to apply systems of commercial reward and punishment in their operations.

C *Conflict-of-interest, potential venality and potent external sources of disruption*

Digital platforms take advantage of privately held user data, as well as more subtle traces left by users. The revelations about Cambridge Analytica's use of personal information are well attested, particularly in its development of a database and algorithm to enable US voters to be influenced at the ballot box through personalised political messaging following the harvesting of millions of Facebook profiles.[44] It was only possible on that scale, tapping into a data set of tens of millions of users, because of the market scale and user following of Facebook. An aspect of the psychological manipulation involved was said to exploit the 'inner demons' of the targets.[45]

Among more recent allegations of voter manipulation is the claim by Robert Epstein, given in evidence in 2020 before the US Senate Judiciary Committee that Google could manipulate 'upwards of 15 million votes'.[46] It was coupled with a recommendation that Google's search index be made public. Google dismissed his research as 'nothing more than a poorly constructed conspiracy theory',[47] and other critics attacked it for lack of peer review and small sample size. The computer science professor, Panagiotis Metaxas, was of the view that, 'I and other researchers who have been auditing search results for years know that this did not happen'.[48] Interestingly, he commented that Epstein's paper demonstrated a possibility of 'what such an influence *could* have been *if* Google was

[43] ACCC Final Report (n 27) 525.
[44] Carole Cadwalladr and Emma Graham-Harrison, 'Revealed: 50 Million Facebook Profiles Harvested for Cambridge Analytica in Major Data Breach' *The Guardian* (online, 18 March 2018) <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>.
[45] Ibid.
[46] Linda Qiu, 'Fact check: Trump Falsely Claims Google "Manipulated" Millions of 2016 Votes' *The New York Times* (online, 19 August 2019) <https://www.nytimes.com/2019/08/19/us/politics/google-votes-election-trump.html>.
[47] Ibid.
[48] Ibid.

manipulating its electoral search results'.[49] No comment is intended here on that evidence beyond noting the *capability* of Google to engage in such manipulation, demonstrating clear need for restraint on Google's part, and (as elaborated further below) human rights oversight and regulation of its activities.

The 2019 report by the Oxford Internet Institute is also instructive in identifying other disruptive forces, keen to manipulate both platforms and their users. The report, entitled *The Global Disinformation Order*,[50] found instances of social media manipulation by governments and political parties ('Cyber troops') in 70 countries, a significant escalation on previous years. Cyber troop activities are not merely part of a blunt authoritarian apparatus to suppress dissent, or (in the case of China) to achieve geopolitical influence but are more widely practised. They include 'political bots' to amplify hate speech (communicating more or less autonomously), micro-targeting, trolling, doxing and harassment.[51] The report concludes that 'computational propaganda has become a ubiquitous and pervasive part of the digital information ecosystem.'[52] Among the messaging strategies used by Cyber troops in conversations with users online are: pro-government or pro-party propaganda; opposition-directed attacks and smear campaigns; distraction and diversion from important issues; fuelling of division and polarisation; and suppressing participation through personal attacks or harassment. The creation of disinformation or manipulated media (often targeted at specific communities) is a common communication strategy (memes, videos, fake news websites or manipulated media) to mislead users. Cyber troops also engage in censorship through the coordinated reporting of unwanted content as 'inappropriate' so that it is taken down. They are said to work in conjunction with private industry, civil society organisations, Internet subcultures, youth groups, hacker collectives, fringe movements, social media influencers, and volunteers who ideologically support their cause.[53] Facebook is the dominant Cyber troop platform not least because of its market size.[54]

If there is ever to be a wholesale review of digital platforms for human rights compliance, and it is suggested there should be, it would have to take account of the reality that they themselves are co-opted for misuse by foreign governments, powerful private political forces, and others. It strengthens the case for State intervention in securing human rights

[49] Ibid.
[50] Samantha Bradshaw and Philip N Howard, 'The Global Disinformation Order: 2019 Global Inventory of Organisation Social Media Manipulation' (Oxford Internet Institute, Computational Propaganda Research Project, 2019).
[51] Ibid 15.
[52] Ibid 2.
[53] Ibid 9.
[54] Ibid 15.

standards in this domain, rather than self-regulation. Any intervention of this sort must itself be suitably transparent and accountable.

The starting point for such a review logically is the digital platforms and their role in regulating content. Given the importance of social media to users as a prevalent contemporary vehicle for exercising their freedom of expression, censorship or manipulation applied by platforms which affects the content of users' communications or what they are able to access, is to be taken extremely seriously. Distortions in what users would otherwise receive (ie but for intervention by digital platforms) when they seek information, or even as it is passively delivered to users, needs to be carefully scrutinised for the operation of censorship, bias or other unwanted influence. The right to hold opinions *without interference* in art 19(1) is, after all, an absolute, unrestricted right.[55] In his latest report, *Disease pandemics and the freedom of opinion and expression*, the UN *Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, David Kaye, astutely recalled the pithy statement by the political philosopher Hannah Arendt, 'How can you have an opinion if you are not informed?'[56] To him, she summed up the theory connecting the right in art 19(1), to hold opinions without interference, with the guarantee in art 19(2), to seek, receive and impart information and ideas of all kinds. He also noted this other comment by Arendt,

> [i]f everybody always lies ... nobody believes anything any longer... And a people that no longer can believe anything cannot make up its mind. It is deprived not only of its capacity to act but also of its capacity to think and to judge. And with such a people you can then do what you please.[57]

In light of growing criticism of digital platforms, their fitness to engage in content moderation and user exclusion under self-devised policies is put into question, particularly while they pursue commercial practices which may cloud their impartiality. As private companies with vested self-interests, they are ill-adjusted for the job of securing neutral human rights observance by users and for users, in a fully transparent and accountable manner.

IV THE QUEST FOR APPROPRIATE CONTENT MODERATION STANDARDS

The obligation rests with those States which are party to the *ICCPR*, such as Australia, to 'respect' *ICCPR* rights (which usually denotes restraint in

---

[55] This is clear from the text of article 19: *ICCPR* (n 10) art 19, but it is further explained in GC 34 (n 20) [9].

[56] Hannah Arendt, 'Hannah Arendt: From an Interview' [1978] (October) *The New York Review*.

[57] Human Rights Committee, *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, UN Doc A/HRC/44/49 (23 April 2020) ('2020 Report') [58], citing 'Hannah Arendt: From an Interview': ibid.

government action), and to 'ensure' those rights (including by regulation of the private sector), to all individuals within its territory and subject to its jurisdiction without distinction of any kind.[58] The obligation to 'respect' restrains the State itself. Restrictions may not extend beyond the permissible limits circumscribed for each right. In the case of freedom of expression, they are set out in art 19(3).[59] The obligation to 'ensure' entails State responsibility to secure adequate protection against restriction from non-State sources, measured by the same permissive limits.[60] While some countries have taken steps to ensure that self-regulation by social media platforms upholds fundamental human rights of users and others,[61] this and related issues have been the focus of recent reports by the Special Rapporteur.

A *The UN Special Rapporteur's stock take*

The Special Rapporteur recently registered his concern at a number of trends that have become apparent in practices involving the use of social media platforms. The first relates to the burgeoning use of social media for conveying hate speech, and inciting violence and discrimination. He emphasised the responsibilities which accompany the commercial operation of social media platforms, to prevent carriage of such material, in spite of the unimaginable scale on which content is added daily in every language. The second is consequential, and paradoxically relates to the content regulation that this necessitates, at least as an immediate practical expedient. The platforms themselves have been best placed, given the stage in which they participate in the sequence of content delivery, and because of their unique technical proficiency, to act as content moderator. This may come at a cost to a platform if it incurs liability for 'publishing' content rather than simply channelling it.[62] Such a censorship function generates a

---

[58] *ICCPR* (n 10) art 2(1).

[59] Human Rights Committee, *General Comment No 31 (80): The nature of the general legal obligation imposed on States Parties to the Covenant: International Covenant on Civil and Political Rights*, 80th sess, UN Doc CCPR/C/21/Rev.1/Add.13 (26 May 2004) [6] ('GC 31').

[60] Ibid [8].

[61] See eg, *Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken or NetzDG* [Network Enforcement Act] (Germany) 1 October 2017, Federal Law Gazette I, 2017, 3352, requires removal of hate speech; France's recent modification to electoral legislation requires removal of 'fake news' during election campaigns. For the position in the UK, see Department for Digital, Culture, Media & Sport, *The Cairncross Review: a sustainable future for journalism* (12 February 2019). In the UK, the government responded in February 2020 to consultation feedback on its 'Online Harms White Paper' describing the government's plans for online safety measures to make companies more responsible for their users' safety online, especially children and other vulnerable groups.

[62] See eg, *Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects on information society services, in particular electronic commerce, in the Internal Market* [2000] OJ L 178/1. The Directive protects intermediaries from liability if operating as a 'mere conduit', 'cache' or 'host'; *Communications Decency Act of 1996*, (United States of America) s 230. See also *Delfi AS v Estonia* (European Court of Human Rights, Grand Chamber, Application No 64569/09, 16 June 2015) (the first case in which the European Court of Human Rights considered the position of the commercial

third, acute, concern extending beyond those identified by the Special Rapporteur. By passing to social media platforms the mantle of content gatekeeper their existing market power and technological power are compounded. The track record of some digital platforms as corporate citizens of ethical good standing is not promising. A censorship role in their hands widens the potential for adverse impact on the conditions in which democratic freedoms are to be enjoyed. The central issue for present purposes is by what standard should content be regulated?

The Special Rapporteur is unequivocal on this point. In his 2018 report to the Human Rights Council,[63] he criticised content regulation by companies by means of so-called 'platform law' in which clarity, consistency, accountability, and remedy were elusive, in spite of the steps taken towards transparency in their rules and government interactions.[64] This led him to propose a framework for the moderation of user-generated online content with human rights occupying centre stage. He also assessed whether States should regulate commercial content moderation and, if so, how. To his mind the guiding universally applicable standard should be that set by art 19 of the Universal Declaration, and its counterpart in binding form, art 19 of the ICCPR which, as the Special Rapporteur put it, contains 'globally established rules' for the protection of freedom of expression.[65] Article 19(3) of the ICCPR encapsulates the parameters by which States Parties are bound when restricting that freedom. Freedom of expression may only be subject to restrictions which are 'provided by law and are necessary for respect of the rights or reputations of others; or for the protection of national security or of public order…or of public health or morals'.[66] In this context the Special Rapporteur drew attention to the *Guiding Principles on Business and Human Rights*, adopted by the Human Rights Council in 2011, establishing in the private sector 'global standard[s] of expected conduct' that should apply throughout company operations and wherever they operate.[67]

Of course, the regulation of freedom of expression also engages duties to ensure enabling environments for freedom of expression and to protect its exercise: '[t]he duty to ensure freedom of expression obligates States to

operator of a professionally managed Internet news portal for offensive comments posted by its readers).

[63] Human Rights Council, *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, UN Doc A/HRC/38/35 (6 April 2018) ('2018 Report').

[64] Ibid [1].

[65] Ibid [5]–[8].

[66] Article 19(3) provides, 'Freedom to manifest one's religion or beliefs may be subject only to such limitations as are prescribed by law and are necessary to protect public safety, order, health, or morals or the fundamental rights and freedoms of others': *ICCPR* (n 10).

[67] Human Rights Council, *Report of the Special Representative of the Secretary-General on the issue of human rights and transnational corporations and other business enterprises*, 17th sess, UN Doc A/HRC/17/31 (21 March 2011) 13, principle 11; 2018 Report (n 63) [10].

promote, *inter alia*, media diversity and independence and access to information.'[68] The importance of media plurality to the Human Rights Committee has already been touched on. The Special Rapporteur for his part, referred to the 2017 *Joint Declaration on Freedom of Expression and 'Fake News', Disinformation and Propaganda* developed by the Special Rapporteur and his counterparts at the Organization for Security and Co-operation in Europe ('OSCE'), the Organization of American States, and the African Commission on Human and Peoples' Rights:

> States have a positive obligation to promote a free, independent and diverse communications environment, including media diversity, which is a key means of addressing disinformation and propaganda. States should put in place other measures to promote media diversity which may include … i. Providing subsidies or other forms of financial or technical support for the production of diverse, quality media content; ii. Rules prohibiting undue concentration of media ownership; and iii. Rules requiring media outlets to be transparent about their ownership structures.[69]

The criteria chosen for content regulation to promote freedom of expression must therefore also be supported by an appropriate enabling environment.

B *Challenges in identifying candidate content for removal*

The content review policies of social media platforms could not be equated with transparency, as it has been a form of self-reporting without rigorous independent verification against clear objective standards. In moderating content, Facebook's policy, for example, has been to remove material that breaches its content policies, reduce the spread of problematic material that does not directly breach those policies, and provide users with additional contextual information about the content that appears in their news feeds.[70] An obvious lack of transparency has existed at another level in the practice of shadow banning, adopted by LinkedIn, Twitter, Facebook and Instagram, where it is not readily apparent, even to the user, that they have been blocked. Facebook and Twitter also down-ranked content or sources in their respective newsfeeds. Facebook has relied extensively on third-party factcheckers in this way:

> If content from a Page or domain is repeatedly given a 'false' rating from our third-party fact-checkers ... we remove their monetization and advertising privileges to cut off financial incentives, and dramatically

[68] 2018 Report (n 63) [6].
[69] Organization for Security and Co-operation in Europe ('OSCE'), *Joint Declaration on Freedom of Expression and 'Fake News', Disinformation and Propaganda* (OSCE, 3 March 2017) 3(a), (d) ('*Joint Declaration*').
[70] Facebook, '*Remove, Reduce, Inform: New Steps to Manage Problematic Content*' (10 April 2019).

> reduce the distribution of all of their Page-level or domain-level content on Facebook.[71]

Facebook's newly established oversight board may be a step in the right direction. It began hearing appeals on 22 October 2020 for blocked or removed content.[72] However, this alone is unlikely to meet the recommendations made by the Special Rapporteur. Among the Special Rapporteur's areas of concern about existing content standards were the vague rules created by digital platforms. He described the rules as 'subjective and unstable bases for content moderation'.[73] Examples of such rules include, Twitter's prohibition of 'behavior that incites fear about a protected group' in its hate, harassment, and abuse policies, and Facebook's distinction between 'direct attacks' on protected characteristics and merely 'distasteful or offensive content'.[74] He pointed to complaints about a tendency for vague hate speech and harassment policies to penalise minorities and reinforce the status of dominant or powerful groups.[75] He therefore stressed the need for platforms to articulate the bases for restrictions on hateful expression, and to demonstrate the necessity and proportionality of any content actions, such as removals or account suspensions, with art 19 in mind.[76] In areas related to news reporting, he acknowledged increasing pressure on platforms to address disinformation spread through links to bogus third-party news articles or websites, fake accounts, deceptive advertisements and the manipulation of search rankings.[77] However, blunt reactions like website blocking or content removal were not the answer, as these risk serious interference with freedom of expression. He voiced particular concern at those measures which enhance restrictions on news content, and the threat they pose to independent and alternative news sources or satirical content.[78]

The Special Rapporteur accepted that the use of automated moderation tools to identify candidate content for removal is necessitated by the vast volume of user generated traffic (as much as 500 hours of video per minute, in May 2020, according to YouTube).[79] While this may suit some applications (like identifying child exploitation material), it is less suitable

---

[71] Oliver Darcy, 'Facebook Touts Fight on Fake News, But Struggles to Explain Why InfoWars Isn't Banned' *CNN Business* (online, 11 July 2018) <https://money.cnn.com/2018/07/11/media/facebook-infowars/index.html>.
[72] See Oversight Board: <https://oversightboard.com/>.
[73] 2018 Report (n 63) [26].
[74] Facebook community standards (hate speech); Twitter rules and policies (hateful conduct policy).
[75] 2018 Report (n 63) [27].
[76] Ibid [26].
[77] Ibid [31].
[78] Ibid.
[79] As of May 2019, more than 500 hours of video were uploaded to YouTube every minute: J Clement, 'Video upload to Youtube every minute 2007-2019', *Statista* (Web page, 25 August 2020) <https://www.statista.com/statistics/259477/hours-of-video-uploaded-to-youtube-every-minute/>.

for others, such as that relating to 'extremism', which requires human evaluation and is context dependent. He touched on the practice of user flagging, which enables anyone to initiate a complaint, based on their own subjective dislike for material, but his concern was the lack of transparency in the selection of flaggers and the standards they apply when removing content. Furthermore, although content moderators exist in large numbers to review flagged content, to adjudicate whether it should be removed (and escalate decision-making for trickier issues), little is known about the processes involved.[80]

Among the consequences for inappropriate content are its removal from an entire platform, a set of platforms, warnings, demonetisation, and (for those responsible for uploading it) account suspension or deactivation.[81] The user who posted it, and the complainant against it, may not even know it has been removed or that other action has been taken, and in those cases when notification is provided it is often couched in generic terms, allowing little understanding by those against whom the action is taken. Transparency reports, where provided, disclose 'the least amount of information about how private rules and mechanisms for self- and co-regulation are formulated and carried out'.[82] The result is that content standards, available only in broad terms, allow platforms discretion without sufficient illumination, even if hypothetical examples and similar are offered by way of explanation.

Available remedies are confined. Crawford and Gillespie described YouTube's system of content determinations some years ago as 'politically closer to a monarchic structure, in the traditional sense of the "rule of one." A plea can be made, but the decision to remove or suspend is solely up to the platform.'[83] Remedies may have improved since then, but those offered by most platforms appeared to the Special Rapporteur to be limited or untimely to the point of non-existence and, in any event, opaque to most users and even civil society experts.[84] The importance of remedies is underscored by the obligation in art 2(3) of the ICCPR on States to ensure that persons whose rights or freedoms have been violated shall have an 'effective remedy', regardless of whether the State or private actors are the source of that violation.[85]

So-called 'community standards' are pervasive and changeable, yet powerfully influential in regulating content. Crawford and Gillespie depict the system of user-initiated complaints and flagging in the following terms:

---

[80] 2018 Report (n 63) [33], [35].
[81] Ibid [36], [37].
[82] Ibid [32]–[40].
[83] Kate Crawford and Tarleton Gillespie, 'What is a Flag For? Social Media Reporting Tools and the Vocabulary of Complaint' (2016) 18(3) *New Media & Society* 410, 422.
[84] 2018 Report (n 63) [38].
[85] See above Part I.

> Engagement with complaints, and the user communities from which they come, is mediated and modulated by the procedures instantiated in the flag interface and the algorithms behind it ... And they are not stable expressions: Their effects are often uncertain and their meaning unclear. They may harken to other democratic and governance processes, but they do not operate within a transparent or representative system … flags are ubiquitous, have a material force in the world, and are indeterminate and steeped in practical politics, yet they have maintained a kind of invisibility from scholarly inquiry.[86]

TV broadcasters routinely apply 'community standards,' in Australia under supervision by the Australian Communications and Media Authority ('ACMA'). However, social media platforms differ from broadcast channels in a number of key respects. The program standards adopted by ACMA pursue particular objectives with a specific notion of community standards in mind. They operate with the aim of developing and reflecting a sense of Australian identity, character and cultural diversity, including through minimum levels of Australian content.[87] There would be little place for such intervention in the regulation of content carried by digital platforms. It is why the Special Rapporteur recommended that platforms incorporate directly into their terms of service and 'community standards' relevant principles of human rights law that ensure content-related actions will be guided by the same standards of legality, necessity and legitimacy that bind State regulation of expression.[88]

Certainly, the expectations are strict when it comes to avoiding bias in a national broadcaster, like the ABC or BBC, and it is important to observe how suspicions of this occur. The sort of conclusions reached in independent reviews of the ABC are that '[m]ost of the complaints about bias have come from the government of the day … both parties have been far less hostile to the ABC when in opposition',[89] and 'when in government the political propensity for political parties is more towards criticism of the ABC than praise.'[90] In response to the British Prime Minister's proposal to remove the BBC's licence fee in reaction to the corporation's election coverage in 2019, the Director-General of the BBC commented that 'the fact criticism came from all sides of the political divide shows to me that we were doing our job without fear or favour.'[91] Whatever may be the truth behind such accusations of systematic bias, they are difficult to make out. However, complaints at specific instances are easier to support, particularly

---

[86] Crawford and Gillespie (n 83) 411.
[87] Australian Communications and Media Authority, *Explanatory Statement – Broadcasting Services (Australian Content) Standard 2016*, 1.
[88] 2018 Report (n 63) [45].
[89] Rhonda Jolly, 'The ABC: An Overview' (Research Paper Series, 2014–15, Social Policy Section, Parliamentary Library) 28–9.
[90] Ibid.
[91] PA Media, 'BBC Chief Dismisses Accusations of Bias in Politics Coverage' *The Guardian* (online, 23 December 2019) <https://www.theguardian.com/media/2019/dec/23/bbc-chief-dismisses-accusations-of-bias-in-politics-coverage>.

in news or public affairs programs carried by the public service broadcaster, where the host indulges excessively in such practices as 'gotcha' questioning, hostility applied unevenly to guests because of their political position, thwarting of their attempts to put their position squarely, questioning in the form of *ad hominem* attack, or treatment otherwise fitting the description of a 'drive by shooting'. Accusations of bias in printed news bear some similarity, but are typically directed at headline grabs which unfairly represent underlying content, failure to carry important stories because of their disadvantage to a particular political cause, and unwarranted slant or spin. Commercial considerations also play a part in content selection.

A further fundamental difference between social media platforms and broadcast media is that the latter has never constituted a rostrum, or other vehicle through which *everyone* is *entitled* to express themselves, a point well made in the Individual opinion of Human Rights Committee member Torkel Opsahl in *Hertzberg et al v Finland*.[92] The same is also true of print media. By contrast, social media platforms do represent a digital soapbox of sorts (even if they are also popular for maintaining contact between family and friends, interacting socially and networking). Digital platforms are generally unconstrained by the codes by which broadcasters are bound (for example by ACMA) because of the particular public objectives they serve. Broadcast codes at least define the community standards which they apply. Embargoing content on social media on the basis of vaguely or subjectively defined 'community standards' allows content to be excluded when much of it should be safeguarded against restriction according to internationally-held human rights standards.

In his 2020 report, the Special Rapporteur tackled a number of conflicting challenges affecting digital platforms in the context of the COVID-19 virus. These underscore the heavy responsibilities they bear. Among them is that platforms are justifiably under pressure to avoid carrying potentially harmful public health disinformation, and he noted that several had taken aggressive steps to address misinformation.[93] Twitter broadened the concept of 'harm' to include 'content that goes directly against guidance from authoritative sources of global and local public health information'.[94] Others are said to have followed an unusually aggressive approach in

---

[92] Individual opinion of Torkel Opsahl in *Hertzberg et al v Finland*, *Communication No 61/1979*, UN Doc CCPR/C/OP/1 (1985) 124 (for him it was decisive that 'nobody – and in particular no State – has any duty under the Covenant to promote publicity for information and ideas of all kinds' and that access to media operated by others is always and necessarily more limited than the general freedom of expression, such that it may be controlled on grounds which do not have to be justified under Article 19(3)).

[93] 2020 Report (n 57) [51].

[94] See Alex Hern, 'Twitter to Remove Harmful Fake News about Coronavirus: Site Changes Rules to Ban Content Aimed at Making People Act Against Official Advice' *The Guardian* (online, 20 March 2020) <https://www.theguardian.com/world/2020/mar/19/twitter-to-remove-harmful-fake-news-about-coronavirus>.

removing misinformation and other exploitative content and boosting trusted content, for example from the World Health Organization.[95] Yet at the same time, the suspension of press briefings which occurred in this crisis frustrates the scrutiny of government which is the basis for proper, informed debate of its actions.[96] There is also the potential to apply curation policies in a discriminatory way. In reference to public protests inconsistent with government guidelines, the Special Rapporteur suggested that platforms should apply policies to all gatherings, without discrimination on the basis of the protesters' viewpoints.[97] He had previously responded to the controversy surrounding Facebook's ban on users organising events that defy government's guidance on social distancing, by remarking that

> [t]he normal democratically accountable way to go about this is that the government issues an order to Facebook to take down particular content through a legal mechanism … This is the difference between accountable governance and unaccountable companies. A government that makes this decision is subject to law; they're subject to real legal constraint and remedy. Facebook is not.[98]

He concluded his survey of some of these tensions by observing that in certain circumstances information saves lives.[99] On the other hand, lies and propaganda deprive individuals of autonomy, of the capacity to think critically, of trust in themselves and in sources of information, and of the right to engage in the kind of debate that improves social conditions. Worst of all, censorship can kill, by design or by negligence.

> These are the principles that have led States, in multiple instruments across human rights law and the political organs of the United Nations, to emphasize government's obligation to enable, promote and protect robust and independent media and provide reliable information to the public, which extends to affirmative government information strategies concerning voting, health and other essential services and fundamental rights.[100]

All this supports a strong measure of scepticism surrounding the appropriateness of digital platforms to exercise the authority they currently do. They do not claim to apply the universal touchstone of art 19 and other clearly stipulated ICCPR standards. Indeed, the above issues raised by the Special Rapporteur indicate areas of departure from that standard. For example, the purpose of Facebook's Oversight Board is to promote 'free expression' by making principled, independent decisions regarding content

---

[95] This and similar issues were canvassed by David Kaye when delivering the Mehmet Ali Birand Lecture: 'Pathogen or repression' (Speech, 3 May 2020).
[96] 2020 Report (n 57) [22].
[97] Ibid [52].
[98] Julia Carrie Wong, 'Facebook Bans Some Anti-lockdown Protest Pages: The Move Raises Thorny Questions About Civil Rights Amid the Coronavirus Pandemic' *The Guardian* (online, 22 April 2020) <https://www.theguardian.com/technology/2020/apr/20/facebook-anti-lockdown-protests-bans>.
[99] 2020 Report (n 57) [99].
[100] Ibid [60].

on Facebook and Instagram and by issuing recommendations on the relevant Content Policy.[101] Terminology such as 'free expression' is capable of meaning different things in different countries, as the earlier discussion on the meaning of 'free speech' in Australia indicates. It does not denote freedom of expression under art 19. Michael Karanicolas has gone so far as to state in the context of the Global South that '[i]n most instances … moderation decisions are not based on any legal standard at all, but on the platforms' own privately drafted community guidelines, which are notoriously vague and difficult to understand.'[102]

C *What if article 19 applied?*

The adoption of art 19 norms would not be popular among digital platforms. It would require rewriting of key algorithms, adjustment from loosely defined policies and other unclear criteria. On the other hand, it would put them, and the States in whose jurisdiction they operate within, in compliance with international human rights law. It would also provide a clear and objective basis for the concessional immunity from liability which they currently enjoy, but which is put at risk if they are regarded as publishing content.

ICCPR standards have the advantage of being based on long-established principles. They operate to exclude material on the basis of particular, well-defined rights. For example, where material: is defamatory or in violation of privacy (art 17, protection against interference with privacy and reputation);[103] portrays child exploitation material (art 24, the rights of the child);[104] constitutes the advocacy of extreme forms of hatred constituting incitement to discrimination, hostility or violence (art 20(2));[105] constitutes a terrorist threat or otherwise jeopardises an individual's personal security (art 6, the right to life, and art 9, the right to security of the person); or subjects individuals to inhuman or degrading treatment (art 7).[106] On each of these issues, the guidance offered by the Human Rights Committee in its respective General Comments directs States on the appropriate content of their domestic laws, so as to produce a high degree of uniformity among

---

[101] Oversight Board (n 72).

[102] Michael Karanicolas, 'Moderate globally, impact locally: A series on content moderation in the Global South' (Information Society Project, Yale Law School, 5 August 2020).

[103] *ICCPR* (n 10) art 17(1) provides: 'No one shall be subjected to arbitrary or unlawful interference with his privacy, family, home or correspondence, nor to unlawful attacks on his honour or reputation'.

[104] Ibid art 24 provides: 'Every child shall have ... the right to such measures of protection as are required by his status as a minor ...'

[105] Ibid art 20(2) provides: 'Any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law'. This supplements more general non-discrimination protection in arts 2 and 26.

[106] Ibid art 6 provides: 'Every human being has the right to life ... No one shall be arbitrarily deprived of his life.', art 7 provides: 'No one shall be subjected to torture or to cruel, inhuman or degrading treatment.' Art 9 provides: 'Everyone has the right to ... security of the person'.

them, even if certain jurisdictional variations may be expected.[107] For ICCPR States Parties, their domestic law should already meet the standards of the ICCPR (subject to any reservations they may have entered upon ratification). Digital platforms may therefore be required to apply the ICCPR-compliant standards of domestic law. Where domestic law departs from ICCPR requirements, as it does in Australia for reasons already discussed, art 19 standards may still be adopted as understood in terms of the detail expressed in General Comment 34. (This approach differs in its emphasis from those measures directed at extreme and obvious forms of harm caused by material that is unlikely to qualify for protection under art 19).[108] The result intended here is a form of regulation of digital platforms which at least recognises and applies protection for freedom of expression and applies criteria for content curation which corresponds with the objective standards of justification for restriction established in art 19(3).[109]

There is a real risk that reliance by social media platforms on their own standards (including home-grown 'community standards') tends to support sensibilities shared by a majority or dominant moiety, at the expense of minorities and the less powerful. Yet, it is the explicit purpose of a number of ICCPR provisions to uphold the freedoms of the ICCPR without discrimination (principally articles 2, and 26 which particularly benefit those who do not stand with the majority),[110] and to uphold particular freedoms for members of minorities, whether religious, political, ethnic or linguistic (articles 18, 19 and 27),[111] who are most likely to suffer at the

---

[107] See General Comments on art 17: *General Comment No 16: Article 17 (The Right to Privacy, family, home and correspondence, and protection of honour and reputation)*, 32nd sess, UN Doc HRI/Gen/1/Rev.9 (Vol I) (8 April 1988); art 24, *General Comment No 17: Article 24 (Rights of the Child)*, 35th sess, (7 April 1989); art 20(2) (GC 34 (n 20)); art 6, *General Comment No 36 Article 6 (Right to Life)*, 124th sess, UN Doc CCPR/C/GC/36 (30 October 2018); art 7, *General Comment No 20: Prohibition of Torture, or Other Cruel, Inhuman or Degrading Treatment or Punishment (Article 7*), 44th sess, UN Doc A/44/40 (10 March 1992); and art 9, *General Comment No 35, Article 9 (Liberty and security of person)*, 112th sess, UN Doc CCPR/C/GC/35 (16 December 2014).

[108] The *Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019* (Cth) requires content, internet, and hosting providers to report abhorrent violent conduct in Australia that is accessible through their services or hosted on their services. This material is unlikely to be protected by art 19, by virtue of *ICCPR* art 5(1) for being aimed at the destruction of *ICCPR* rights and freedoms.

[109] *ICCPR* (n 10) art 19(3).

[110] Article 2(1) requires States to respect and ensure *ICCPR* rights 'without distinction of any kind' on stated grounds. Article 26 requires that 'the law shall prohibit any discrimination and guarantee ... equal and effective protection against discrimination' on stated grounds: Ibid art 2(1).

[111] Ibid arts 18, 19, 27. Article 18 provides, 'Everyone shall have the right to freedom of thought, conscience and religion.' Article 27 provides that 'persons belonging to [ethnic, religious or linguistic minorities] shall not be denied the right ... to enjoy their own culture, to profess and practise their own religion, or to use their own language.' The guidance on these provisions is found in *General Comment No 22: Article 18 (Freedom of Thought, Conscience or Religion)*, 48th sess, UN Doc CCPR/C/21/Rev.1/Add.4 (30 July 1993), and

hands of prevailing social mores. A truly pluralistic media provides channels for representing the diversity of opinion that exists within any society. If digital platforms hold a finger on the scales, intentionally or unintentionally, in the transmission of news feeds, or if they exercise bias of any sort in the removal of content or user access to a platform, the human rights cost could be far greater than to the individual users directly affected. A tilted platform fundamentally upsets the presuppositions of a democratic society on which the UN human rights system, and Australian society, is founded.

Content exclusion and tagging by platforms are currently predicated on notions such as 'fake news', 'hate' and 'dangerous misinformation', notoriously vague notions, with little to support the necessity of the interference as required by art 19(3). Such crude appliqués provide the pretext for peremptory content loss, or user bans, which are difficult to challenge. The principal concern is that these crucial matters should not be determined by the policies of private companies, still less, algorithms and those who operate their human faculties in distinguishing truth from misinformation according to subjective, non-transparent standards.

As to how matters might be remedied to that end, the starting point, as suggested by the Special Rapporteur, is that content-related restrictions should meet the standards of legality, necessity and legitimacy set by article 19.[112] This should produce less dependence on vague terminology currently applied when excluding content. The principle of legality stems from article 19(3) and requires restrictions on freedom of expression to be 'provided by law'.[113] The law 'must provide sufficient guidance to those charged with [its] execution to enable them to ascertain what sorts of expression are properly restricted and what sorts are not'.[114] Restrictions with a basis only at common law (rather than in legislation) will suffice, provided they satisfy those criteria.[115] Mere policy is generally not enough.[116] The rub in Australia comes with upholding freedom of expression, including the right to access information, as a right of the individual (when the implied right of political communication is *not* a personal right),[117] assertible against the State as guarantor bound by the terms of limitation in art 19(3). The Human Rights Committee's General

Comment No 23: Article 27 (Rights of Minorities), 50th sess, UN Doc CCPR/C/21/Rev.1/Add.5 (8 April 1994).

[112] 2018 Report (n 63) [45].

[113] *ICCPR* (n 10) art 19(3).

[114] GC 34 (n 20) [25].

[115] *Gauthier v Canada*, CCPR/C/65/D/633/1995, 5 May 1999 [13.5]; *Dissanayake v Sri Lanka*, CCPR/C/93/D/1373/2005, 22 July 2008 [8.2].

[116] Netherlands (Antilles) A/37/40 (1982) 109 (the objection to guidelines issued by the Prime Minister on the freedom of civil servants to express their opinions outside the civil service in broad and general terms was that any restrictions should be laid down by law).

[117] *Australian Capital Television v Commonwealth* (1992) 177 CLR 106, 150; *Lange v Australian Broadcasting Corporation* (1997) 189 CLR 520, 560.

Comment 31 stresses the need to observe terms of limitation such as found in art 19(3), yet art 19(3) has no counterpart which operates substantively in Australian domestic law.[118] It would appear that legislation would therefore be needed in Australia to incorporate art 19 standards in this context. Its focus would be the curation function of digital platforms. It would be practically arduous for public authorities to undertake that task, but it is inevitable, given the criticisms of digital platforms already made, that platforms need to be closely scrutinised. Public authorities charged with that supervisory responsibility must also be fully accountable. The repetitive findings of infringement of competition law suggest that digital platforms may not be readily amenable to compliance, as long as profit gains outweigh fines. Options for such proposals should take account of that. Further elaboration is beyond the scope of this piece, which merely makes the case for this issue to be reviewed in light of Australia's international obligations. As already noted, some countries have already taken steps to ensure digital platforms uphold fundamental human rights of users and others, and Australia should too.

Cardinal to any standardised system for moderating freedom of expression (as with other human rights) must be clarity, predictability, transparency, and accountability. Clarity in the priority of upholding the freedom. Predictability in the precise terms on which it is guaranteed and may be permissibly restricted, especially in the principles then applied (legality, necessity and proportionality). Transparency in the specific rules that correspond to those criteria, and in their operation (this requires notification and clear explanation of incidents of content removal or user disqualification). Accountability by appeal, ultimately by court processes, against all forms of censorship, and remedies where it is unwarranted, or for abuse of entrusted responsibility. The stipulations are all met in article 19 compliance.

V CONCLUSION

Although social media offer a powerful podium for exercising freedom of expression by individuals, it is clearer now than ever before that more is required of the private digital platforms to uphold universal standards of freedom of expression, including the constituent right to seek and receive information.

The use of digital platforms, including for expressing and accessing information and opinion, and news delivery, demonstrates the need for human rights protection as a matter of State obligation. By default, rather than design digital platforms have assumed responsibility to protect their users in an environment in which an unstoppable tide is generated of genuine hate speech, incitement to violence, glorification of terrorism, and

[118] GC 31 (n 59) [6].

child exploitation material, among other harmful content. It is not surprising that the burden for addressing this has rested with those with the technical capability, combined with the contractual capacity, to regulate users. In the tradition of Euripidean Greek tragedy, the digital platforms appeared spontaneously as the *deus ex machina* to resolve this disorder.[119]

However, it is necessary that the platforms themselves operate in a manner which upholds the human rights of all users. The most obvious and universally applicable criteria for curating content are found in article 19 of the ICCPR. These are the standards which the Human Rights Committee applies when monitoring ICCPR compliance across all issues affecting freedom of expression, including when ensuring plurality of the media.[120] They are the very standards strongly recommended by the Special Rapporteur.[121] However, article 19 has not yet been the yardstick chosen by the platforms. A focus on users, vested interests and other preferences has resulted in the selection of other criteria, including self-determined 'community standards.'

The process of content curation is best undertaken, or at least supervised closely, by a non-participant if the process is to be objective. The obvious candidate is the State. Australia is already obligated under the ICCPR, even if it has not implemented the ICCPR's terms in domestic law, and this article urges that it should assume an active role in the regulation and supervision of digital platforms.

The power of influence available to digital platforms, whether by human or algorithmic operation, is largely imperceptible to the user. It may occur in the news mix that is selected for presentation in the processes that render it 'trustworthy' or fact-checked; in the content that is not seen because it is downgraded; in the criteria applied to the surfacing of information, and in the removal of particular content or de-platforming of individuals through enforcement of content policies. Monetisation and other privileges operate as the currency of further unseen preferences affecting the operations of digital platforms. The power imbalance in favour of the technocracy and lack of transparency which would allow it to be challenged is like a two-way mirror. This is a matter of State obligation to monitor and where necessary correct. Since freedom of expression is not absolute (though freedom of opinion is), what are the rules to be applied to order the relationships which exist in the operation of digital platforms? Surely the politically neutral standards of article 19 and other provisions of the ICCPR. The duty to ensure freedom of expression burdens States with the task of promoting, among other things, media diversity and independence,

[119] Literally, 'A God from a machine,' it is referring in dramatic context to a surprise event which saves the day, against all odds.
[120] See above IIIA.
[121] See above IV.

and free access to information.[122] These are particularly important today, as innumerable issues of profound importance face governments in every country, and societies are deeply divided on the best solutions to be pursued in particular national contexts.

[122] Joint Declaration, 3(a). See also GC 34 (n 20) [18], [40].