6-2019

Singapore's latest efforts at regulating online hate speech

Siyuan CHEN
Singapore Management University, siyuanchen@smu.edu.sg

Chen Wei CHIA

SINGAPORE'S LATEST EFFORTS AT REGULATING ONLINE HATE SPEECH: A PERSPECTIVE FROM INTERNATIONAL LAW AND INTERNATIONAL PRACTICES

CHEN SIYUAN* AND CHIA CHEN WEI**

The introduction of the Protection from Online Falsehoods and Manipulation Act (POFMA) has been generating considerable debate and feedback. Some of the concerns raised include whether the bill unduly restricts the freedom of expression. In focusing on the hate speech provisions of the POFMA, this legislation comment situates the criticisms within the larger framework of international human rights law and international practices and proposes some ways forward to improve the regulatory framework for online hate speech.

I. SETTING THE CONTEXT

1. The internet may have been around for more than a couple of decades, but governments around the world continue to find it a great challenge to regulate online speech in a way that strikes an appropriate balance between freedom and security[1] and that also accords with jurisdictional propriety.[2] The latest conundrum presented by the world wide web is supposedly that of "fake news", which is claimed to have the potential of disrupting society by sowing discord and division between groups, influencing political outcomes, legitimising fringe views, and discrediting establishment journalism.[3] Given the nature of the internet, the spreading of "fake news" is not confined to persons within a single jurisdiction, and can include persons from other jurisdictions and even bots, through coordinated means or otherwise.

2. In other countries, responses by governments to this phenomenon have ranged from proposed fines[4] and imprisonment[5] to actual shutdowns of internet services.[6] In

* *LLB (First Class Hons) (National University of Singapore), LLM (Harvard); Associate Professor, Singapore Management University*. Contact: siyuanchen@smu.edu.sg

** *LLB (Magna Cum Laude) (Singapore Management University); Justices' Law Clerk, Supreme Court of Singapore.* This article is written in the author's personal capacity, and the opinions expressed here are entirely his own.

[1] See generally UN 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 (2018).

[2] See generally Chia Chen Wei, "Sketching the Margins of a Borderless World" (2018) 30(3) Singapore Academy of Law Journal 833.

[3] See generally Cherilyn Ireton and Julie Posetti, *Journalism, Fake News & Disinformation* (UNESCO: 2018).

[4] See for instance BBC, "Websites to be Fined Over "Online Harms" Under New Proposals" (8 April 2019): https://www.channelnewsasia.com/news/asia/sri-lanka-social-media-shutdown-raises-fears-on-free-expression-11469136.

[5] See for instance Saudi Gazette, "5-Year Jail, SR3m Fine for Social Media Material that Disrupt Public Order" (9 May 2018): http://saudigazette.com.sa/article/542723/SAUDI-ARABIA/5-year-jail-SR3m-fine-for-social-media-material-that-disrupt-public-order.

[6] See for instance CNA, "Sri Lanka Social Media Shutdown Raises Fears on Free Expression" (23 April 2019): https://www.channelnewsasia.com/news/asia/sri-lanka-social-media-shutdown-raises-fears-on-free-expression-11469136.

Singapore, we have the recent introduction of the Protection from Online Falsehoods and Manipulation Act (now popularly known as POFMA),[7] a fairly substantial piece of legislation that comprises 62 sections.[8] The POFMA has been generating considerable debate and feedback since it was first mooted in 2017, be it in the form of the protracted Select Committee hearings on Deliberate Online Falsehoods,[9] the Green Paper on Deliberate Online Falsehoods,[10] or the public consultations that have ensued after the First Reading of the bill.[11] International groups such as the International Commission of Jurists,[12] Reporters Sans Frontiers,[13] and Human Rights Watch[14] have also weighed in on the matter, claiming that the law would unduly suppress speech and possibly be abused by the government.

3. The objective of this legislation comment is threefold: first, to identify and analyse some of the issues presented by the critical provisions of the POFMA, with a particular focus on the online hate speech provisions; secondly, to situate these hate speech provisions within the framework of both international law and international practices; and thirdly, to make some brief recommendations on how else the law can be improved. To be clear, the POFMA's ambit extends beyond online hate speech, but those areas will not form the focus on this comment due to space constraints.

II. EXAMINING THE POFMA PROVISIONS AND IDENTIFYING SOME OF THE POSSIBLE CRITICISMS

4. We begin our analysis by examining the salient POFMA provisions that pertain to online hate speech, before setting out some of the potential issues:

    a. Under section 7(1), a person must not do any act in or outside Singapore to communicate[15] a statement knowing or having reason to believe[16] that it is a false statement of fact, and the communication of that statement is likely to,

[7] 10/2019.

[8] The government has also stated that efforts in improving media and information literacy and promoting fact-checking would be taken as well: The Straits Times, "Parliament: Law Against Online Falsehoods Will Not Stifle Free Speech, Say Ministers" (1 April 2019): https://www.straitstimes.com/politics/parliament-law-against-online-falsehoods-will-not-stifle-speech-ministers. Indeed, simply developing an appetite for longform debates would go some way in neutralising simplified soundbites on, say, power structures and identity politics that seed the ground for online hate speech.

[9] The result was a 317-page report: 13th Parliament of Singapore, "Report of the Select Committee on Deliberate Online Falsehoods", Parl 15 of 2018.

[10] Misc 10 of 2018.

[11] See also Ministry of Law, "New Bill to Protect Society from Online Falsehoods and Malicious Actors" (1 April 2019): https://www.mlaw.gov.sg/content/minlaw/en/news/press-releases/New-Bill-to-Protect-society-from-Online-Falsehoods-and-Malicious-Actors.html.

[12] International Commission of Jurists, "Singapore: ICJ Calls on Government Not to Adopt Online Regulation Bill in Current Form" (12 April 2019): https://www.icj.org/singapore-icj-calls-on-government-not-to-adopt-online-regulation-bill-in-current-form/.

[13] Reports Without Borders, "RSF Explains Why Singapore's Anti-Fake News Bill is Terrible" (8 April 2019): https://rsf.org/en/news/rsf-explains-why-singapores-anti-fake-news-bill-terrible.

[14] Human Rights Watch, "Singapore: Reject Sweeping "Fake News" Bill" (3 April 2019): https://www.hrw.org/news/2019/04/03/singapore-reject-sweeping-fake-news-bill.

[15] This is defined in section 3(2) as a statement that is made available to one or more end-users in Singapore on or through the internet, MMS, or SMS.

[16] *Cf* Sui Yi Siong *et al*, "Written Representation 130 to the Select Committee" (7 March 2018) at para 7.

inter alia, prejudice national security,[17] public health, public safety, or public tranquillity, or incite feelings of enmity, hatred, or ill-will between different groups of persons.[18] As we will soon make clear and define, the latter half of this provision (especially the part on incitement of feelings) essentially refers to the consequence of hate speech, while the former half refers to the act that causes it (a false statement of fact).[19]

b. Under section 7(2), a person guilty of section 7(1) shall be fined not exceeding $50,000 or be imprisoned for a term not exceeding 5 years, or both. If the person had used an inauthentic online account or a bot to commit the offence, he shall be fined not exceeding $100,00 or be imprisoned for a term not exceeding 10 years, or both.

c. Under sections 10, 11, and 12, a minister may order a correction direction or a stop communication direction if a false statement of fact has been communicated in Singapore and the minister is of the opinion that it is in the public interest to issue the direction. "In the public interest" is defined non-exhaustively in section 4, and the definition that is of relevance here is the doing of something that is necessary or expedient: in the interest of the security of Singapore; to protect public health, public safety, or public tranquillity; or to prevent incitement of feelings of enmity, hatred, or ill-will between different groups of persons.

d. Under section 15, a person who fails to comply with an order made under sections 11 or 12 without reasonable excuse shall be fined not exceeding $20,000 or be imprisoned for a term not exceeding 12 months, or both.

e. Under sections 17 and 19, a person subject to an order made under sections 11 or 12 may appeal to the High Court after he has applied to the minister to vary or cancel the order, and the minister has refused the application in whole or in part.[20]

f. Under sections 20, 21, 22, and 23, a minister may order an internet intermediary to make a general correction, make a targeted correction, or disable access if a false statement of fact has been communicated in Singapore and the minister is of the opinion that it is in the public interest to issue the direction.[21]

---

[17] *Cf* section 3(1)(e) of the Sedition Act (Cap 290, 2013 Rev Ed): "A seditious tendency is a tendency … to promote feelings of ill-will and hostility between different races or classes of the population of Singapore."

[18] *Cf* section 8(1)(a) of the Maintenance of Religious Harmony Act (Cap 167A, 2001 Rev Ed): "The Minister may make a restraining order … where the Minister is satisfied that that person has committed or is attempting to commit … acts … causing feelings of enmity, hatred, ill-will or hostility between different religious groups".

[19] In recent times, the link between the two has manifested in its most violent form in debates on immigration policy and refugee policy: UN Human Rights Council, "Report of the Independent International Fact-Finding Mission on Myanmar", UN Doc A/HRC/39/64 (2018).

[20] See also Siraj Omar, "Protection from Online Falsehoods and Manipulation Bill: A More Calibrated Approach", *The Straits Times* (1 May 2019): "[The POFMA] does not provide a deadline within which the minister must decide. This should be expressly set out in subsidiary legislation (as is usually the case), and should ideally be kept short so as to enable the aggrieved person to have recourse to the courts without undue delay." The Ministry of Law later said it would consider this suggestion: K Shanmugam, "NMPs Agree on Major Points of Falsehoods Bill", *The Straits Times* (3 May 2019).

[21] See also the government powers under the Broadcasting Act (Cap 28, 2012 Rev Ed).

g. Under section 27, a person who fails to comply with an order made under sections 21, 22, or 23 without reasonable excuse shall be fined not exceeding $20,000 or imprisoned for a term not exceeding 12 months, or both; in any other case, the fine shall not exceed $1 million.

h. Under section 29, a person subject to an order under sections 21, 22, or 23 may appeal to the High Court after he has applied to the minister to vary or cancel the order, and the minister has refused the application in whole or in part.

i. Under section 57, any offence under the statute may be compounded, while under section 61, the minister may exempt any person or class of persons from any provision of the statute.

5. Before going further, it is important to note that the Explanatory Statement and Preamble of the POFMA explicitly state the various purposes[22] of the legislation. What is of relevance here are the stated aims of preventing the electronic communication of false statements of fact and information manipulation as well as the enabling of measures to counteract the effects of such communication.[23] The Ministry of Law has also clarified that: the bill targets falsehoods and not opinions, criticisms, satires, and parodies; the primary remedial measure would be corrections, and not removal or takedown of content; criminal offences would apply only to malicious actors seeking to undermine society; and any decision by the government over what is false can be overridden by the courts on appeal.[24] These positions were reiterated in Parliament during the debate over the bill.[25]

6. What then are some of the issues that arise from the provisions that have been set out above? Here, we identify no less than four issues that have been raised by various parties; these also correspond to the chief concerns that were raised in Parliament during the debate over the legislation.[26] The first and perhaps most obvious one is that concerning definitions, the chief example of which is what constitutes a "false statement of fact" under section 7. Section 7 does not define the phrase,[27] but the definitional clause (section 2) does clarify that the determination of a statement of fact is based on a "reasonable person" analysis and that false statements would include partly or wholly

[22] It is trite law in Singapore that the purposive interpretation trumps all other canons of statutory interpretation: section 9A of the Interpretation Act (Cap 1, 2002 Rev Ed).
[23] See also The Straits Times, "PM Lee Hsien Loong: Legislation Essential to Curbing Spread of Fake News" (26 April 2019): https://www.straitstimes.com/politics/pm-legislation-essential-to-curbing-spread-of-fake-news, where it was suggested that the POFMA can be applied in a preventive way that focuses on the intent to do harm, without harm necessarily materialising.
[24] Ministry of Law, "New Bill to Protect Society from Online Falsehoods and Malicious Actors" (1 April 2019): https://www.mlaw.gov.sg/content/minlaw/en/news/press-releases/New-Bill-to-Protect-society-from-Online-Falsehoods-and-Malicious-Actors.html. See also The Straits Times, "No Need to be Overly Worried about Fake News Laws, Says Ong Ye Kung" (29 April 2019): https://www.straitstimes.com/singapore/no-need-to-be-overly-worried-about-fake-news-laws-says-ong.
[25] Singapore Parliament Debates, 8 May 2019.
[26] The Straits Times, "Parliament: Shanmugam Addresses 5 Concerns Over Proposed Fake News Law" (8 May 2019): https://www.straitstimes.com/politics/parliament-law-minister-shanmugam-addresses-five-concerns-over-proposed-fake-news-law.
[27] Indeed, the ambiguities present in the POFMA extends even to terms that have been defined – "in the public interest" being the prominent example – the definition seems circular when we compare the language between sections 4 and 10 (or 11, 12, 20, 21, 22, and 23 for the matter).

misleading statements.[28] Notwithstanding this attempt at disambiguation, it is not always easy to distinguish between facts and opinions, characterisations, and misimpressions that result from decontextualisation.[29]

7. To illustrate, when a fatal riot broke out in Charlottesville during the Unite the Right rally, President Donald Trump was deemed by his political opponents to have committed hate speech by supporting white supremacists when he said that "You had some very bad people in the group … But you also had people that were very fine people, on both sides".[30] Given the context of the rally – in that contrary to the media's characterisation, many who protested the removal of Confederate monuments were not necessarily "alt-right fascists" – the President's words could be interpreted a number of ways (and even as a fact), but this opinion (of his support for neo-Nazis) has since been presented as fact, for instance, by various Democratic presidential candidates for the 2020 campaign.[31]

8. Even scientific claims, which are supposed to be independent of political ideologies, are not spared of this uncertainty. Consider the example of biological sex and gender. From a neuroscience perspective, there is now growing evidence that there are generally differences between male and female brains that account for, say, dissimilarities in behaviour, cognitive functions, and rates of mental disorders.[32] However, academic studies looking to investigate these differences have readily been condemned by postmodernists and intersectionalists for downplaying the existence of structural inequalities and worse, as factual non-starters, misogynistic, and hateful.[33] This may sound like caricature, but it is not, and even supposed leading institutions like Cambridge University have fallen prey to this line of thinking.[34]

9. The illustrations we have cited highlight the importance of needing the POFMA to strike the right balance between allowing internet discourse on matters of public interest to flourish (maybe accepting that mistakes on the facts necessarily have to be made before a consensus can be reached) and ensuring that misunderstandings that lead to hate speech and possibly violence are minimised.[35] This brings us to the second issue, which concerns the allocation of powers in the fact-determination process. The POFMA leaves no doubt that it is the government (minister) who determines whether a false

---

[28] Though in a different context, article 16 of the General Data Protection Regulation (2019) Directive 95/46/EC has shown that the concept of incomplete accounts of information is difficult to ascertain and adjudicate upon.

[29] See for instance *Salov v Ukraine* App No 65518/01 (ECtHR, 6 September 2005) at para 113.

[30] ABC News, "Trump Said "Blame on Both Sides" in Charlottesville, Now the Anniversary Puts Him on the Spot" (12 August 2018): https://abcnews.go.com/Politics/trump-blame-sides-charlottesville-now-anniversary-puts-spot/story?id=57141612.

[31] See for instance YouTube, "Joe Biden for President" (25 April 2019): https://www.youtube.com/watch?v=VbOU2fTg6cI. One could, of course, attempt to justify Biden's speech as political and therefore more deserving of protection (see for instance *Stankov and the United Macedonian Organisation Ilinden v Bulgaria* App nos 29221/95, 29225/95 (ECtHR, 2 October 2001) at para 102), but that notion does not have universal acceptance – not least in Singapore laws, including the POFMA.

[32] Stanford Medicine, "Two Minds" (2017): https://stanmed.stanford.edu/2017spring/how-mens-and-womens-brains-are-different.html.

[33] Quillette, "Denying the Neuroscience of Sex Differences" (29 March 2019): https://quillette.com/2019/03/29/denying-the-neuroscience-of-sex-differences/.

[34] Quillette, "Cambridge Capitulates to the Mob and Fires a Young Scholar" (2 May 2019): https://quillette.com/2019/05/02/cambridge-capitulates-to-the-mob-and-fires-a-young-scholar/.

[35] *Cf* Siraj Omar, "Protection from Online Falsehoods and Manipulation Bill: A More Calibrated Approach", *The Straits Times* (1 May 2019).

statement of fact has been made, and whether it is in the public interest to order correction or stop directions, general or targeted corrections, or disabling of access. But when it comes to online hate speech regulation, are there superior alternatives?

10. The usual rejoinder to letting the government decide is to let either the intermediary or the online community self-regulate, though often it is a combination of both. For instance, Facebook and Twitter have in place user term agreements and community standards, reporting and flagging mechanisms, moderators, algorithmic filters, and inhouse legal teams to track potential hate speech – in many jurisdictions, they seldom wait for a government order before acting (to remove the post, suspend the user, and so forth).[36]

11. The problem though is their abject lack of transparency whenever intermediaries moderate content or suspend or ban accounts, leading some to conclude that intermediaries have the proclivity to censor based on their own political ideologies, often seeing no difference between facts and hate speech – Facebook and Twitter are routinely accused of considering contentious social issues like abortion and transgenderism as not up for debate (even when factual claims are made) because of the supposed potential for generating hate speech, but there are neither appeal mechanisms nor reasons given when they limit or block such speech.[37] Seen in this light, putting aside the difficulty of determining what is hate speech, maybe it is actually more democratic and accountable for the government, rather than intermediaries, to make the call on when the freedom of expression should be limited. We should add briefly that having the courts make the call is not a viable alternative, for reasons that the government have highlighted during the debate: courts are not designed to act quickly, both in terms of getting a hearing date and allowing the full presentation of evidence in adversarial system of law.[38]

12. But even if one accepts that the government is best placed to make the call on whether something qualifies as hate speech, something might be said about the obstacles to challenging the government's decision. Once the minister decides that directions, corrections, or disablement are appropriate, these potentially draconian orders then must be complied with until an application to vary or cancel them succeeds, or potentially heavy sanctions could follow (this is elaborated in the next issue). Further, though the minister's decision to refuse variation or cancellation is appealable to the courts, this necessarily involves litigation, which entails time, resources, and costs,[39] though it was clarified in Parliament that the appeal process will be simple (no court fees for the first three days) and fast (as little as nine days for the court to hear the case).[40]

---

[36] This will be elaborated in the next section.

[37] USA Today, "Is Facebook Too Liberal?" (3 May 2018): https://www.usatoday.com/story/tech/news/2018/05/03/facebook-pledges-investigate-charges-bias-against-conservatives/574505002/; CNBC, "Why Silicon Valley Tech Giants Can't Shake Accusations of Anticonservative Political Bias" (17 October 2018): https://www.cnbc.com/2018/10/17/why-silicon-valley-cant-shake-accusations-of-anticonservative-bias.html.

[38] See Singapore Parliament Debates, 8 May 2019.

[39] See also Siraj Omar, "Protection from Online Falsehoods and Manipulation Bill: A More Calibrated Approach", *The Straits Times* (1 May 2019). There is, of course, also uncertainty as to how the courts will respond to such applications.

[40] The Straits Times, "Appeal Process Under Fake News Law Will Be Simple, Fast" (8 May 2019): https://www.straitstimes.com/singapore/appeal-process-under-fake-news-law-will-be-simple-fast-shanmugam.

13. One could of course rationalise this on the basis that hate speech is irredeemably bad for society, and so it stands to reason to make it hard for the maker of the speech or the intermediary hosting the speech to challenge the government's decision; further, as much as the government is responsible for upholding rights, it also has the responsibility to protect competing rights, such as the safety and security of vulnerable groups. Yet this presupposes that – apart from its ease of determination – such speech has no inherent value whatsoever, and the best remedy is governmental intervention. In the final analysis though, once it is accepted that the government is better to place to determine the legality of content, there would invariably be transactional costs involved to challenge it. How these costs can be meaningfully managed may well go beyond the issue of costs of litigation in this specific context, but the relatively high costs of litigation in Singapore in general.

14. The last issue we have identified is the proportionality of the consequences once the minister has decided that there is hate speech. Directions to stop and orders to disable access are not extraordinary and are indeed logical responses once one accepts that hate speech warrants legislative and executive action, but corrections might be tricky to implement. For individuals, the maker of the statement is expected to declare that his statement is false and point to where the truth is to be found on terms to be decided by the government. For intermediaries, by virtue of them being passive entities even though they are often passive entities that do not generate (but sometimes can control) user content, they have to communicate corrections on terms to be decided by the government. Because it is the government that decides the extent of corrections required, for the reasons stated above, this might be preferable to letting the intermediaries decide.

15. Having said that, in the context of hate speech, it seems unlikely in most cases that the government would opt for corrections – stop or disablement orders would more likely be preferred.[41] One would also imagine that in Singapore, hate speech, especially that pertaining to race or religion, is a presumptively egregious offence.[42] A person found guilty of such speech could be fined up to $50,000 and imprisoned for up to 5 years; if he refuses to abide by any government order to stop or modify his statement, he could be subject to further sanctions in the form of a fine up to $20,000 and imprisonment of up to 12 months. The fining of an intermediary of up to $1 million may not seem much if we have in mind the revenues of social media giants such as Facebook and Twitter, but this amount is for each breach and not a global cap.

16. All things considered, it seems there is, at multiple junctures, a perceptible potential for chilling speech and self-censorship, but as earlier stated, the purpose of this comment is to compare the POFMA with international norms, or at least those that are identifiable. This give us a better idea as to whether the POFMA under- or over-regulates what has clearly become a societal menace, and also provides some yardsticks on what the best practices could be in combatting online hate speech – after all, the

[41] Even then, this potentially places intermediaries that operate in multiple jurisdictions between a rock and a hard place – while the POFMA requires them to comply with directions as there is no defence of duty under law, other jurisdictions may sanction them for unduly restricting the freedom of expression of the person who posted the content.

[42] See section 3(1)(e) of the Sedition Act (Cap 290, 2013 Rev Ed); section 8(1)(a) of the Maintenance of Religious Harmony Act (Cap 167A, 2001 Rev Ed) (both of which have been set out earlier).

validity of any criticism of the POFMA is best seen in the light of international human rights law and practices, for any other point of reference would not have the same pedigree of international consensus. Additionally, the view that effective internet governance requires intergovernmental collaboration in formulating harmonised rules is increasingly gaining traction.[43] It is perhaps for these reasons that the Ministry of Law was minded to include an annex on international developments when introducing the POFMA,[44] as was the Select Committee in its report.[45] With all that said, how does the POFMA hold up under international legal scrutiny?

III. THE INTERNATIONAL HUMAN RIGHTS FRAMEWORK FOR REGULATING ONLINE HATE SPEECH

17. Any discourse on international human rights, especially where speech is concerned, must invariably begin with the International Covenant on Civil and Political Rights (ICCPR)[46] – the preeminent multilateral human rights treaty[47] that has been ratified by more than 170 of the 193 UN Member States (more will be said soon about Singapore not being a state party). What must be observed at the outset is that although the freedom of expression is one of the fundamental rights protected by the covenant, this freedom[48] is not absolute and can be limited by states.[49] Article 19(3) of the ICCPR permits the restriction of the freedom of expression on the grounds of respect of the rights or reputations of others or the protection of national security, public order, public health, or morals.[50] The same provision states, however, that states in interfering with this right must show that any such restriction is provided by law and is necessary; while the threshold for prescription by law is low and gives states some latitude to frame laws broadly, the element of necessity requires the pursuit of a legitimate aim, the existence of a pressing social need, and a proportionate response that has no less restrictive alternatives (the three-part test).[51]

---

[43] See for instance UN, "Global Cooperation and Regulation Key in Addressing Multi-layered Threats Posed by New Technology" (14 November 2018): https://www.un.org/sustainabledevelopment/blog/2018/11/global-cooperation-and-regulation-key-in-addressing-multilayered-threats-posed-by-new-technology/.

[44] Ministry of Law, "New Bill to Protect Society from Online Falsehoods and Malicious Actors" (1 April 2019): https://www.mlaw.gov.sg/content/minlaw/en/news/press-releases/New-Bill-to-Protect-society-from-Online-Falsehoods-and-Malicious-Actors.html.

[45] 13th Parliament of Singapore, "Report of the Select Committee on Deliberate Online Falsehoods", Parl 15 of 2018 at Annexes A–F.

[46] 999 UNTS 171 (entry into force 23 March 1976).

[47] The International Bill of Rights comprises the ICCPR, the International Covenant on Economic, Social and Cultural Rights, 993 UNTS 3 (entry into force 3 January 1976) and the Universal Declaration of Human Rights, 10 December 1948, 217A(III).

[48] Article 19(2) states that everyone "shall have the right to freedom of expression; this right shall include freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice."

[49] *Cf* article 19(1), which states that everyone "shall have the right to hold opinions without interference."

[50] Article 18(3), which pertains to the contiguous right of freedom of religion, similarly restricts religious expressions that might cause harm to public safety, order, health, morals, or the fundamental rights and freedoms of others. See also principle 6 of The Johannesburg Principles on National Security, Freedom of Expression and Access to Information (1995): "expression may be punished as a threat to national security only if a government can demonstrate that: (a) the expression is intended to incite imminent violence; (b) it is likely to incite such violence; and (c) there is a direct and immediate connection between the expression and the likelihood or occurrence of such violence."

[51] This has been the unanimous position across all international human rights bodies, be it the UN (see for instance *Corinna Horvath v Australia* UN Doc CCPR/C/110/D/1885/2009 (HRC, 27 March 2014) at para 3.11), the

18. Then there is article 20(2), which unlike article 19 that requires the state to protect the freedom of expression, requires the state to prohibit certain kinds of speech. Specifically, it obligates states to prohibit speech that advocates national, racial, or religious hatred that constitutes incitement to discrimination, hostility, or violence.[52] This has been referred to by the UN High Commissioner for Human Rights and various UN Special Rapporteurs as the hate speech prohibition, and has been confirmed to apply to online speech as well.[53]

19. In determining whether there is hate speech, factors to be considered include the context in which the speech was made, the status of the maker of the speech, the intent of the maker, the content and form of the speech, the reach of the speech, and the likelihood and imminence of harm (inchoate acts are therefore included).[54] Although it has been said that any restrictions should only be imposed on "the most severe and deeply felt form of opprobrium"[55] and must also fulfil the aforementioned prescription and necessity requirements,[56] it is widely accepted that the ICCPR conception of the limits on freedom of expression is different from, say, the US conception, which sets the high watermark for permissible speech.[57] Recent UN bodies have even claimed that the approach towards online hate speech should be "zero tolerance", in unequivocal opposition to any ideal of relying heavily on the marketplace of ideas to counteract hate speech.[58]

20. Operating in the backdrop is the doctrine of margin of appreciation or margin of discretion, which claims that because national authorities are best placed to balance conflicting fundamental human rights based on each state's unique social context and

---

Strasbourg court (see for instance *Avram v Moldova* App no 41588/05 (ECtHR, 5 July 2011) at para 24), the Inter-American court (see for instance *Tristán Donoso v Panama*, Preliminary Objection, Merits, Reparations, and Costs Judgment (IACtHR, 27 January 2009) at para 56), or the African Human Rights court (see for instance *Interights v Mauritania* AHRLR 87 Comm no 242/2001 (ACommHPR, 2004) at paras 78–79). See also UN Human Rights Committee, "General Comment 34" (12 September 2011) UN Doc CCPR/C/GC/34 at paras 22–34.

[52] See also article 4 of the International Convention on the Elimination of All Forms of Racial Discrimination, 660 UNTS 195 (entry into force 4 January 1969): "States Parties condemn all propaganda and all organisations which are based on ideas or theories of superiority of one race or group of persons of one colour or ethnic origin, or which attempt to justify or promote racial hatred and discrimination in any form, and undertake to adopt immediate and positive measures designed to eradicate all incitement to, or acts of, such discrimination".

[53] UN Human Rights Council, "Rabat Plan of Action on the Prohibition of Advocacy of National, Racial or Religious Hatred That Constitutes Incitement to Discrimination, Hostility or Violence", UN Doc A/HRC/22/17/Add 4 (2013) at para 29.

[54] UN Human Rights Council, "Rabat Plan of Action on the Prohibition of Advocacy of National, Racial or Religious Hatred That Constitutes Incitement to Discrimination, Hostility or Violence", UN Doc A/HRC/22/17/Add 4 (2013) at para 29.

[55] UN Human Rights Council, "Rabat Plan of Action on the Prohibition of Advocacy of National, Racial or Religious Hatred That Constitutes Incitement to Discrimination, Hostility or Violence", UN Doc A/HRC/22/17/Add 4 (2013) at para 29.

[56] UN 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 (2018) at para 8.

[57] This has been the case since the SCOTUS decision interpreting the Amendment I of the Constitution of the US in *Brandenburg v Ohio*, 395 US 444 (1969), which held that the mere advocacy is not enough to constitute hate speech; there must be circumstances that create a clear and present danger that will bring about substantive evils that the government has a right to prevent. This decision has not been altered in spite of developments in internet speech.

[58] See for instance UN, "Third Committee Experts Warn Racism, Hate Speech, White Supremacy to Become Mainstream Unless States Enforce Zero-Tolerance Policies, Prevent Exclusion" (29 October 2018): https://www.un.org/press/en/2018/gashc4245.doc.htm.

circumstances, there is no one-size-fits-all solution when assessing the justifiability of interferences with human rights. However, there are obstacles to concluding that this doctrine enjoys unqualified universal support.

21. For instance, while the European Court of Human Rights (ECtHR) has consistently affirmed it over decades of its jurisprudence,[59] the UN body which monitors the implementation of the ICCPR has been more tentative in its embracement,[60] as has numerous UN Special Rapporteurs.[61] From this standpoint, any invocation of the margin of appreciation outside the context of the European Convention on Human Rights[62] should, in principle at least, be done with some caution.[63] Instead, absent customary international law to the contrary, the text of the ICCPR should remain the first port of call when determining the scope of the freedom of expression, the interpretation of which is of course guided by the rules of treaty interpretation set out in the Vienna Convention on the Law of Treaties.[64]

22. The above establishes the contours for regulating freedom of expression in general. Things become considerably trickier when internet websites enter the framework. For a start, the ICCPR was originally conceived to regulate affairs between the state and its people rather than the state and media entities, let alone intermediaries such as social media websites, the chief purveyors of virulent content;[65] further, before the advent of the internet, the freedom of expression was largely exercised on platforms fundamentally different from the most popular types of media today in terms of barriers to entry, costs, editorial intervention, speed of dissemination, reach, the potential for virality, privacy and anonymity, and the likelihood of fabrication and decontextualisation.[66]

23. The nett result is that in the context of speech made on the internet, it becomes unclear who should be responsible for content regulation such as the curbing of hate speech – is it still the state, as was the case before online freedom of expression was possible, or do internet websites, and in particular social media companies, now bear the main responsibility? And even if the allocation of responsibility can be determined, as we have alluded to above, what exactly should be done to remedy the breach, and what consequences should follow if the breach is not remedied?

---

[59] See for instance *Palomo Sánchez v Spain* App nos 28955/06, 28957/06, 28959/06, 28964/06 (ECtHR, 12 September 2011) at para 54; *Bédat v Switzerland* App no 56925/08 (ECtHR, 29 March 2016) at para 54.

[60] See for instance *Ilmari Länsman v Finland* UN Doc CCPR/C/52/D/511/1992 (HRC, 14 October 1993) at para 9.4.

[61] See for instance UN Human Rights Council, "Promotion and Protection of the Right to Freedom of Opinion and Expression", UN Doc A/71/150 (2016).

[62] ETS 5 (4 November 1950). Having said that, there is no material difference between article 19 of the ICCPR and its equivalent in the ECHR in terms of the grounds in which the freedom of expression may be limited.

[63] The European cases that have consistently upheld that speech that offends and shocks the conscience is permissible (see for instance *Sürek v Turkey* App no 23927/94 (ECtHR, 8 July 1999) at para 58), should, likewise, be treated with some caution.

[64] 1155 UNTS 331 (entry into force 27 January 1980). Typical canons of interpretation, as set out in article 31, include the interpretation of good faith in accordance with the ordinary meaning of the words in their context and in the light of the relevant treaty's object and purpose.

[65] See UN 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/32/38 (2016) at para 9.

[66] See Iginio Gagliardone *et al*, *Countering Online Hate Speech* (UNESCO: 2015) at pp 13–15.

24. In this regard, we can look at how online hate speech and intermediaries have been regulated in various parts of the world. It used to be the case that the importance of self-regulation, as well as the need to avoid broad and sweeping restrictions on internet content, were acknowledged almost universally – by the UN Special Rapporteur on Freedom of Opinion and Expression, the Organisation for Security and Co-operation in Europe, the Organisation for American States, the African Commission on Human and Peoples' Rights, to name but a few regional and international bodies.[67]

25. It also used to be the case that there was near-universal recognition that because the internet has become the principal means by which people exercise their freedom of expression, attempts to restrict online speech on matters of public interest should be condemned.[68] Because of this, the default position in many jurisdictions was that even though they often had billions in revenue, the responsibilities imposed on intermediaries were either unarticulated or minimal, and speech (including hate speech) could only be lawfully restricted upon the issuance of a court order or executive order from the authorities.[69]

26. Complementing this was the intermediaries' repeated assurances that less restrictive alternatives of speech regulation worked more effectively to curtail hate speech – apart from the aforementioned user term agreements and community standards, reporting and flagging mechanisms, moderators, and algorithmic filters to put them on notice of problematic content, in more extreme cases intermediaries could also limit access, suspend accounts, block accounts, demonetise accounts, remove posts, or reduce the visibility of trending posts.[70] Indeed, apart from copyright infringement and child pornography – for which both can be detected with success rates with the right software – there was no expectation that intermediaries had to do much with respect to content regulation.[71]

27. However, giving intermediaries a free pass is no longer the state of affairs. The ECtHR, for instance, now applies various factors for the determination of intermediary liability when an intermediary is perceived to be ineffectual and inefficient in policing content amounting to hate speech.[72] While the court recognises that value judgments, opinions, potentially defamatory speech, and crude expressions with a low register of style are still permissible, where an intermediary allows for content amounting to "manifest expressions of hatred" to remain on its platform, it would be held accountable for being

---

[67] See OSCE, "Joint Declaration on Freedom of Expression and the Internet" (1 June 2011): https://www.osce.org/fom/78309.

[68] UN 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/17/27 (2011) at para 58.

[69] UN 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/17/27 (2011) at para 43; Inter-American Commission on Human Rights, "Freedom of Expression and the Internet" OEA/SER L/II CIDH/RELE/INF 11/13 (2013) at para 138.

[70] See generally Rebecca MacKinnon *et al*, *Fostering Freedom Online* (UNESCO: 2014); UN Human Rights Council, "Report of the Special Rapporteur on Minority Issues" UN Doc A/HRC/28/64 (2015).

[71] See generally Daphne Keller, "Internet Platforms" (2018) Hoover Institution Aegis Series Paper No 1807.

[72] See for instance *Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt* v Hungary App no 22947/13 (ECtHR, 2 May 2016) at para 69. The factors include the context of the comments, the measures applied by the intermediary company to prevent or remove defamatory comments, the liability of the actual authors of the comments as an alternative to the intermediary's liability, and the consequences of the domestic proceedings for the applicant company. For the (similar) position under the European Court of Justice, see *Google France, Google Inc v Louis Vuitton Malletier SA* C-236/08 (CJEU, 23 March 2010) at para 120.

asleep at the wheel, especially if it has the means to determine the legality of posts.[73] In part, this was a rebuke to the notion that the marketplace of ideas would work well enough on the internet, especially since intermediaries are often guilty of perpetuating echo chambers by controlling the content that users interact with based on the users' political preferences – any previously held assumptions about counter-narratives neutralising bad speech were thus shattered.[74]

28. At first, intermediaries resisted taking on this responsibility in monitoring content. The claim was that there was too much content on social media to monitor, and it was unreasonable to impose any duty to regulate when they had neither creative nor editorial oversight for information that could be published by anyone at anytime and anywhere – not to mention the perennial issue of the difficulty in determining hate speech in different contexts.[75] Facebook, for instance, continually insisted for years that it was ill-placed to tackle the problem of fake news facilitating hate speech being shared on its platform before grudgingly accepting that it had to implement measures to fact-check news articles – only to be accused of introducing political bias into such checks.[76] For Twitter, it was only in late-2018 that it revised its hate speech policy,[77] while YouTube continues to be criticised for not being proactive enough in taking down videos that encourage or depict racial and religious violence.[78] But increasing waves of national legislation have followed the ECtHR's lead, even enhancing the burden that is to be placed on intermediaries.

29. The most notable example is probably Germany, which has grappled with immigrant and refugee assimilation in recent years. It passed the notorious Network Enforcement Act in 2017. Under that law, absent any specific agreement with law enforcement authorities, social network intermediaries are to remove "manifestly unlawful" content within 24 hours of receiving a user complaint, as opposed to executive or judicial notice; this time period is only extended to 7 days for unlawful content that is not "manifestly unlawful".[79] Within this framework, intermediaries are expected to consult their own lawyers in determining whether something would count as illegal hate speech and the swiftness of the takedown required – failure to comply with the law may lead

[73] *Delfi AS v Estonia* App no 64569/09 (ECtHR, 16 June 2015) at paras 142–156. In that case, the contents in question were left online for no less than six weeks. See also OSCE, "Joint Declaration on Freedom of Expression and "Fake News", Disinformation and Propaganda", (3 March 2017): https://www.osce.org/fom/302796.

[74] John Samples, "Why the Government Should not Regulate Content Moderation of Social Media" (2019) Cato Institute Policy Analysis 865.

[75] *Cf* The Economist, "Mark Zuckerberg Says He Wants More Regulation for Facebook" (6 April 2019): https://www.economist.com/business/2019/04/06/mark-zuckerberg-says-he-wants-more-regulation-for-facebook.

[76] See for instance Mark Zuckerberg, "Status Update – 13 November 2016" (13 November 2016): https://www.facebook.com/zuck/posts/10103253901916271; Jonathan Vanian, "Facebook CEO Mark Zuckerberg Admits "Huge Mistake" But Will Not Step Down" (4 April 2018): http://fortune.com/2018/04/04/facebook-mark-zuckerberg-data-cambridge-analytica/.

[77] Wired, "Twitter Releases New Policy on "Dehumanising Speech"" (25 September 2018): https://www.wired.com/story/twitter-dehumanizing-speech-policy/.

[78] See for instance Bloomberg, "YouTube Executives Ignored Warnings, Letting Toxic Videos Run Rampant" (2 April 2019): https://www.bloomberg.com/news/features/2019-04-02/youtube-executives-ignored-warnings-letting-toxic-videos-run-rampant. Active monitoring is, of course, only one step away from pre-emptive censorship and prior restraint.

[79] The Act to Improve Enforcement of the Law in Social Networks (12 July 2017): https://www.bmjv.de/SharedDocs/Gesetzgebungsverfahren/Dokumente/NetzDG_engl.pdf?__blob=publicationFile&v=2.

to fines of up to 50 million Euros,[80] a far cry from the hundreds of Euros the ECtHR is used to sanctioning for online hate speech.[81]

30. Even France[82] and the UK,[83] long the bastions of *laïcité* and the right to offend respectively, want to adopt something similar in the face of rising anti-Semitism and Islamophobia expressed online. So too in countries outside Europe, such as India.[84] And in the wake of the Christchurch shootings, Australia plans to fine intermediaries up to 10% of the platform's annual turnover if they do not remove violent material expeditiously.[85] Indeed, so-called rights-conscious jurisdictions now have, or will soon have laws that are more severe as compared to jurisdictions such as Malaysia (where the maximum fine is around USD 120,000)[86] and Russia (where the maximum fine is around USD 15,000).[87]

31. Lest it be assumed that the situation is somehow different for liability for individuals – it is not. It is no longer uncommon for jurisdictions to provide for fines and imprisonment sentences to punish those that knowingly spread fake news and disinformation that are tantamount to hate speech. States such as Bangladesh,[88] Egypt,[89] and Malaysia,[90] which mete out both considerably hefty fines and imprisonment sentences to tackle the so-called spread of disinformation that compromises national stability and security, are no longer anomalous. In addition to Germany,[91] UK[92] also

---

[80] Taylor Wessing, "Germany's Network Enforcement Act and its Impact on Social Networks" (6 August 2018): https://www.lexology.com/library/detail.aspx?g=fb107efe-70ae-4e97-9913-5035aeeb518a.

[81] See for instance *Delfi AS v Estonia* App no 40287/98 (ECtHR, 16 June 2015) at paras 27–31. See also Reuters, "Turkey Fines Twitter for Failure to Remove "Terrorist Propaganda": Official" (11 December 2015): http://www.reuters.com/article/us-turkey-twitter-fine-idUSKBN0TU0NK20151211.

[82] Politico, "Macron Vows Measures to Tackle Online Hate Speech and Anti-Semitism", (21 February 2019): https://www.politico.eu/article/macron-vows-measures-to-tackle-online-hate-speech-and-anti-semitism/. France already has laws dealing with fake news, but this is limited to the spread of misinformation during elections: Assemblee Nationale, "Dispositions Modifiant le Code Electoral" (20 November 2018): http://www.assemblee-nationale.fr/15/ta/tap0190.pdf.

[83] HM Government, "Online Harms White Paper" (April 2019): https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/793360/Online_Harms_White_Paper.pdf.

[84] Forbes, "India's New Rules to Govern Social Media Raise Fears of More Censorship" (22 January 2019): https://www.forbes.com/sites/meghabahree/2019/01/22/indias-new-rules-to-govern-social-media-raise-fears-of-more-censorship/#5ed7cc106759.

[85] Attorney-General for Australia, "Tough New Laws to Protect Australians from Live-Streaming of Violent Crimes" (4 April 2019): https://www.attorneygeneral.gov.au/Media/Pages/Tough-New-Laws-to-protect-Australians-from-Live-Streaming-of-Violent-Crimes.aspx.

[86] CLJ Law, "Anti-Fake News Bill 2018": https://www.cljlaw.com/files/bills/pdf/2018/MY_FS_BIL_2018_06.pdf.

[87] CNN, "New Law Lets Russia Jail People Who "Disrespect" the Government Online" (7 March 2019): https://edition.cnn.com/2019/03/07/europe/russia-internet-law-intl/index.html.

[88] Forbes, "Bangladeshi Digital Security Act Draws Fire From EU" (28 September 2018): https://www.forbes.com/sites/emmawoollacott/2018/09/28/bangladeshi-digital-security-act-draws-fire-from-eu/#1e1df4a60277.

[89] Associated Press, "Egypt Tightens Restrictions on Media, Social Networks" (20 March 2019): https://www.apnews.com/1540f1133267485db356db1e58db985b.

[90] CLJ Law, "Anti-Fake News Bill 2018": https://www.cljlaw.com/files/bills/pdf/2018/MY_FS_BIL_2018_06.pdf.

[91] DW, "German Court Sentences Facebook User to Jail for Xenophobic Comments" (17 October 2016): https://www.dw.com/en/german-court-sentences-facebook-user-to-jail-for-xenophobic-comments/a-36069082.

[92] HM Government, "Online Harms White Paper" (April 2019): https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/793360/Online_Harms_White_Paper.pdf.

plans to imprison persons guilty of online disinformation campaigns, following the lead by the likes of Canada,[93] Italy,[94] and Kenya.[95] These states clearly do not think intermediary responsibility diminishes individual responsibility.

IV. WHERE SINGAPORE STANDS AND OTHER RECOMMENDATIONS TO IMPROVE OUR REGULATORY FRAMEWORK

32. If there was once a time that the exercise of freedom of expression online was to be freely celebrated for giving everyone a potentially powerful voice, that time has in our view passed. Whereas previous forms of ground-breaking mediums – be it the microphone, radio, broadcast television, or the internet – were eventually found not to have fundamentally altered the landscape for freedom of expression such as to warrant paradigm shifts in regulation, states now consider social media to be the genuine game-changer (for the worse).

33. This is reflected most obviously in how the interpretation of the ICCPR has changed drastically in the form of the new and emerging online hate speech laws we have just surveyed above; the internet and social media may be ubiquitous, but it has not resulted in a unified understanding of how best to regulate the problem both in terms of offence creation and the punishments that follow. Whereas the previous emphasis was on invoking article 19(3) of the ICCPR only in extreme scenarios, states have now in effect pivoted to using article 20(2) as the default starting point, complemented by the use of the margin of appreciation and contiguous international human rights obligations such as those found in ICERD.[96] The Rabat Plan used for determining hate speech is also decidedly open-ended enough without drawing any bright lines – much will depend on the facts of each case and the circumstances of each jurisdiction, a situation no doubt complicated by the amplifying nature of social media. The composite picture that emerges from all of this is not one that is terribly reassuring for those looking to discern discrete rules of international law in this domain – and one has not even begun to discuss if a clear line can be drawn between incitement laws and anti-blasphemy laws, the latter of which has been condemned by the UN[97] but can, in certain situations where statements about race or religion are made, be hardly distinguishable from the former.

34. As far as Singapore is concerned, it has not ratified the ICCPR, and so it is not bound by it as a matter of treaty law.[98] But given the very high ratification numbers,[99] putting

[93] Postmedia, "Convicted Hate-Monger Gets Added Jail Time for his Muslim-Offending "Social Experiment" (6 January 2015): https://o.canada.com/news/blatchford-convicted-hate-mongerer-gets-added-jail-time-for-his-muslim-offending-social-experiment.

[94] Washington Post, "A Man was Sentenced to 9 months in Prison. His Crime? Posting Fake Reviews on TripAdvisor" (12 September 2018): https://www.washingtonpost.com/world/2018/09/12/man-tried-sell-fake-tripadvisor-reviews-hes-going-prison-after-landmark-ruling-italy/?utm_term=.1ccc1d46566d.

[95] NPR, "Kenya's Crackdown on Fake News Raises Questions About Press Freedom" (19 May 2018): https://www.npr.org/sections/thetwo-way/2018/05/19/612649393/kenyas-crackdown-on-fake-news-raises-questions-about-press-freedom.

[96] See footnote 52.

[97] See UN Human Rights Office of the High Commissioner, "Elimination of All Forms of Religious Intolerance", UN Doc A/72/365.

[98] Singapore also subscribes to the dualist approach to the domestic incorporation of international law: see generally Chen Siyuan, "The Relationship between International Law and Domestic Law" (2011) 23(1) SAcLJ 350.

[99] The 19 states that have neither signed nor ratified do not include any major states.

aside issues of enforcement,[100] there is some force in the claim that articles 19 and 20 of the ICCPR have attained the status of customary international law.[101] Yet even assuming the two articles bind Singapore, the aforementioned three-part test for justifying interferences with article 19 can probably be surmounted with some ease.

35. Specifically, vague or broad as some of the POFMA provisions may be, the threshold for prescription for law has always been low and almost never successfully argued before international tribunals, and in any event, to the extent that the prescription requirement can be satisfied with safeguards,[102] the minister's decision can be judicially reviewed and would constitute such safeguards.[103] On the next requirement, few would, or can, quibble with the suppression of online hate speech as fulfilling a pressing social need and constituting a legitimate aim, notwithstanding the aforesaid difficulties in identifying what might qualify as hate speech.[104] But, also as mentioned, the Rabat Plan looks at multiple factors such that what may be hate speech in one place may not be hate speech in another – the factors of likelihood of harm, for instance, is going to be so context-sensitive. When viewed through an international lens that often minimises the review of domestic measures, this is not going to be an insurmountable hurdle. Lastly, the proportionality of the POFMA sanctions, be it for individuals or intermediaries, would likely be situated in the middle of the spectrum of punishments for online hate speech when compared to current and developing international norms across the board.[105] And if the margin of appreciation does apply (at any given limb of the three-part test), this will only bolster Singapore's position, considering its political history with respect to managing racial and religious relations.[106]

36. Having said that, apart from the four issues we had previously highlighted and analysed, we do think the regulatory framework can be improved in the following ways in the light of what has been developing internationally. After all, given that international developments are now veering more towards the Singapore-style of governance for regulating online content, this is a rare opportunity for us to lead the way in setting standards. First, the current tenor of the POFMA may give the impression that, unlike the case for individual offenders who may already face punishment for posting hate speech online, intermediaries are better off waiting for the government to issue an order before acting to remove hate speech hosted on their platforms instead of adopting a more proactive stance in monitoring hate speech.

---

[100] See UN Human Rights Office of the High Commissioner, "Human Rights Bodies – Complaints Procedures": https://www.ohchr.org/en/hrbodies/tbpetitions/pages/hrtbpetitions.aspx.

[101] The requisite elements are that of widespread and representative state practice and *opinio juris*: *North Sea Continental Shelf Cases*, ICJ Reports 1969 at paras 73–77.

[102] *Liu v Russia (no 2)* App no 29157/09 (ECtHR, 26 July 2011) at para 88.

[103] *Cf* UN 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 (2018) at para 23.

[104] See generally *Robert Faurisson v France* UN Doc CCPR/C/58/D/550/1993 (HRC, 2 January 1993).

[105] Further, the assessment of whether the requirements of prescription by law, the existence of a pressing social need, and a proportionate response that has no less restrictive measures are met, must necessarily be conducted against the backdrop of the specific circumstances of a state and take into account the "specific need on which they are predicated": UN Human Rights Committee, "General Comment 34" (12 September 2011) UN Doc CCPR/C/GC/34 at para 22.

[106] See also ECtHR, "Hate Speech Fact Sheet" (March 2019): https://www.echr.coe.int/documents/fs_hate_speech_eng.pdf.

37. While we have argued – as has the UN recently[107] – that the government is better placed than intermediaries to determine what is online hate speech and that a government order is probably the most democratic option when it comes to this type of censorship,[108] this should not leave the tools that intermediaries already have (filters, complaint mechanisms, moderators to enforce community standards, and so forth) without teeth. There would be situations in which intermediaries would, through their own mechanisms or otherwise, be put on notice about potentially problematic content – for instance, in the wake of a terrorist attack, or on the eve of elections. When this happens, intermediaries should work in concert with the authorities to decide on the best course of action, rather than revert to private censorship or do nothing at all. Preferably, however, any decision made this way should have some element of public access so that a hybrid form of private-public censorship does not ensue. If we look for example at Facebook's recent mass-bans of American and British public figures, their decisions have not been based on facts but simply their own feelings and political preferences[109] – something which should not happen (at least not so easily) under the POFMA.

38. Our second point is related to this: transparency in any decision-making remains important for public confidence in the system to be maintained. Indeed, a commentator has suggested that the POFMA should "require a Minister's order to identify the relevant falsehood, set out what the true position is, identify the specific public interest involved and how it is threatened by the falsehood, and articulate why the order is both proportionate and necessary" and also "expressly require any order to be proportionate to the nature of the falsehood and the degree of harm to the public interest."[110]

39. The same commentator has also suggested that as a matter of practice, the government "should provide an annual summary of the Ministerial orders issued, the facts and circumstances of each case, the reasons for the specific Ministerial orders, the number of appeals … and the outcome of the appeals … This annual review would enable Parliament to decide if the law is properly achieving its stated legislative aims".[111] Adopting these two suggestions does incur some costs and is not strictly required even by international standards, but we are broadly in agreement with the suggestions insofar as they promote transparency and Singapore continues to pride itself on being one of the most transparent countries in the world.[112]

---

[107] See generally UN 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 (2018).

[108] See also UN 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/32/38 (2016) at paras 51–55.

[109] Reuters, "Facebook Bans Alex Jones, Other Extremist Figures" (3 May 2019): https://www.reuters.com/article/us-facebook-extremists-usa/facebook-bans-alex-jones-other-extremist-figures-idUSKCN1S82D7. A common tactic they use is to brand groups they disagree with as "extremist" or "far-right". As private corporations, they can claim to be not obligated to uphold constitutional rights (in the form of freedom of expression), but they should not be given free rein to do as they please.

[110] Harpreet Singh, "Strengthening the Online Falsehoods Bill: Some Practical Suggestions" (April 2019) Singapore Law Watch at paras 9 and 13. *Cf* Siraj Omar, "Protection from Online Falsehoods and Manipulation Bill: A More Calibrated Approach", *The Straits Times* (1 May 2019).

[111] Harpreet Singh, "Strengthening the Online Falsehoods Bill: Some Practical Suggestions" (April 2019) Singapore Law Watch at para 21. See also Straits Times, "Nominated MPs to Propose Four Amendments to Fake News Bill" (1 May 2019): https://www.straitstimes.com/politics/nominated-mps-propose-four-amendments-to-fake-news-bill.

[112] The Straits Times, "Singapore Rises to Third Place in Annual Ranking of Least Corrupt Countries" (29 January 2019): https://www.straitstimes.com/politics/singapore-rises-to-third-place-in-annual-ranking-of-least-corrupt-countries-global-study.

40. We do not think, however, as some commentators have suggested, that an additional layer of checks should exist in the form of an independent body.[113] Expediency is key to the efficient (and cost-effective) operation of the POFMA, and a check already exists in the form of the courts. What is of greater importance is the accountability aspect. While the Ministry of Law has stated that the government's reasons for each order will be guided by subsidiary legislation, it may be more prudent to build this into the primary legislation. Subsidiary legislation takes time to develop and does not have the same binding effect as primary legislation, and as pointed out by several Nominated Members of Parliament during the debate, can be more easily amended (in a negative sense) than primary legislation.[114] Given the uncertainties over what would constitute online hate speech, requiring reasons to be given from the very first orders that are issued pursuant to the POFMA would provide important certainty and strengthen the rule of law. This certainly would help not just with regard to what constitutes an offence, but also why certain orders were made and considered proportionate (which is separate from the proportionality of the sanctions that flow from violating the orders).

41. Finally, we foresee difficulties in justifying the inclusion of messaging services as intermediaries that would fall under the POFMA (whether in relation to online hate speech or otherwise). Under section 2, one of the examples given for what counts as an intermediary is "internet-based messaging services"; section 3 also confirms that material is considered to be communicated for the purposes of the POFMA so long as it is sent through the internet. However, messaging services are meant to facilitate private communications. In contrast, the whole point of posting on social media is to hope that the content goes viral – which is hard to achieve unless there is some publicity of the content, even if it is only being circulated within closed groups. Another difference between a messaging service and social media is that of reach. Even before recent changes to limit the forwarding of messages,[115] apps like WhatsApp and Telegram just do not have the same design architecture and features as Facebook or Twitter to facilitate sharing of messages with the same scale.

42. More importantly, there is also the issue of privacy. In pretty much any given case, how is one supposed to monitor what is being shared between users of messaging services without some sort of surveillance?[116] This could explain why in countries where messaging services have been regulated for online hate speech, the governmental response has seldom deviated from a complete shutdown of those services.[117] When

[113] See The Straits Times, "Shanmugam Responds to Key Issues on Fake News Bill" (4 May 2019): https://www.straitstimes.com/politics/law-minister-responds-to-key-issues-on-fake-news-bill. It was also stated in Parliament that such a body would not have the same democratic mandate as the government to make decisions: The Straits Times, "Why Independent Council to Review Government's Actions Is Not Necessary" (8 May 2019): https://www.straitstimes.com/politics/why-independent-council-to-review-govts-actions-is-not-necessary.

[114] See Singapore Parliament Debates, 8 May 2019.

[115] CNN, "WhatsApp Tightens Limit on the Number of People You Can Share Messages With" (21 January 2019): https://edition.cnn.com/2019/01/21/tech/whatsapp-forwarding-limits-india/index.html.

[116] As to whether there are privacy rights in Singapore, see Chen Siyuan, "The Regulatory Framework for Aerial Imaging by Recreational Users of "Drones" in Singapore: Old and Emerging Issues and Some Possible Solutions" (2017) 29(1) Singapore Academy of Law Journal 126.

[117] See for instance Business Insider, "The Turkish Government Reportedly Blocked WhatsApp and Other Social Media Sites" (4 November 2016): https://www.businessinsider.com/social-media-and-messaging-sites-blocked-in-turkey-2016-11/?IR=T; The Straits Times, "Indonesian Government Lifts Ban on Telegram" (12 August 2017): https://www.straitstimes.com/asia/se-asia/indonesian-government-lifts-ban-on-telegram; CNA, "Sri Lanka Social Media Shutdown Raises Fears on Free Expression" (23 April 2019):

this was raised in Parliament, it was not stated how the government would be alerted of problematic content being shared through such services. Instead, the Senior Minister of State said that "in closed spaces, people are more susceptible to emotive falsehoods, because these are the spaces inhabited by the familiar and the trusted."[118] Of course, one of the ways to work around this is to have the messaging service prevent the sharing of certain websites or certain online articles. But this reintroduces the problems of private censorship, which on balance, is probably not desirable, even if the censorship is limited to reducing visibility and is based on user feedback.[119] Ultimately, in deciding to include messaging services, one should bear in mind that the pursuit of security still needs to be tempered by some sense of proportionality and efficacy. This challenge may well be made more difficult as "deep fakes" become more prevalent, but that is something we would leave to for a future endeavour.

https://www.channelnewsasia.com/news/asia/sri-lanka-social-media-shutdown-raises-fears-on-free-expression-11469136.

[118] The Straits Times, "Closed, Encrypted Communications Also Covered by Law" (8 May 2019): https://www.straitstimes.com/politics/closed-encrypted-communications-also-covered-by-law-edwin-tong.

[119] See also The Straits Times, "Facebook Rolls Out Fact-Checking Service in Singapore to Combat Fake News" (3 May 2019): https://www.straitstimes.com/tech/facebook-rolls-out-fact-checking-service-in-singapore-to-combat-fake-news; The Daily Wire, "Journalism Schools Poynter Publishes List of "Unreliable News"" (2 May 2019): https://www.dailywire.com/news/46703/journalism-school-poynter-publishes-list-ashe-schow.