Response to the Online Safety Reform Legislation Consultation

by Dr Kim Barker and Dr Olga Jurasz

February 2020

Dr Kim Barker
Stirling Law School
University of Stirling
Scotland
kimberley.barker@stir.ac.uk

Dr Olga Jurasz
Open University Law School
Open University
England
olga.jurasz@open.ac.uk

We are responding to the call for evidence in our capacity as experts on social media abuse, online violence against women, online misogyny, and internet regulation. We have in the past made significant contributions to the UN calls for evidence on online violence against women, to the Bracadale Review on Hate Crime in Scotland, the UK House of Commons Women and Equalities Committee inquiry into sexual harassment of women and girls in public spaces, and to Scottish Government's 'One Scotland: Hate Has no Home Here' Consultation on amending Scottish hate crime legislation.[1] In addition, we have made representations to the Scottish Government as to the need to amend legislation to cover a wider range of harassing and abusive behaviours online. We have recently published a world-leading volume '*Online Misogyny as a Hate Crime: a Challenge for Legal Regulation*' (Routledge 2019). We have been working on issues relating to harassment of women and girls in online spaces since 2013. We are possibly your only evidence respondents that have experience of the wider issues surrounding online harassment, and who take a holistic approach to the legal problems posed by such harassment, merging criminal law, gender, human rights, and internet law expertise. We would add that we are happy to provide further expertise or evidence if this is of use. We are only commenting on the questions posed from the perspective of our research, which generally focuses on online abusive behaviours, offensive content, internet regulation (broadly conceived) and the impact abusive content has on the participation and safety of women online. This expertise is placed within considerations of legal responses to addressing the challenges posed by illegal, harmful and abusive online content, with specific focus on the gender dimension of these pressing issues. This submission builds upon our response to the Australian Government Online Safety Charter response from April 2019.[2]

[1] Kim Barker and Olga Jurasz, 'Submission of Evidence on Online Violence Against Women to the UN Special Rapporteur on Violence Against Women, its Causes and Consequences, Dr Dubravka Šimonović' (Open University, November 2017) http://oro.open.ac.uk/52611/; Kim Barker and Olga Jurasz, 'Submission of Evidence to Scottish Government Independent Review of Hate Crime Legislation (Bracadale Review)' (Open University, December 2017) http://oro.open.ac.uk/52612/; Kim Barker and Olga Jurasz, 'Written Submission of Evidence to the Women and Equalities Committee Inquiry into Sexual Harassment of Women and Girls in Public Spaces' (Open University, March 2018) http://oro.open.ac.uk/53804/.

[2] Kim Barker and Olga Jurasz, 'Australian Government Online Safety Charter Response'(April 2019) https://www.communications.gov.au/have-your-say/online-safety-charter-consultation.

Summary

We welcome the Australian Government's objectives for reform and proposals for addressing problematic content online through law. Whilst the objectives outlined in the discussion paper are praiseworthy and ambitious, we have concerns about the scope and the umbrella notion of 'online safety' as well as 'one size fits all' approach evident in the document and in the Online Safety Reform Agenda more broadly.[3]

We recognize and accept that making the Internet a safer place for children and young people has been prioritised not only in Australia but also in other jurisdictions – notably, Germany, the UK, and New Zealand[4] but note that there is a disproportionate level of attention paid within the Paper and the proposals to services marketed to and targeted at children. Given the disproportionate level of attention focused on these users, there are other elements of online safety that are conspicuous by their absence from the reform proposals. In particular, the absence of text-based abuse[5] is a notable omission within the legislative and policy proposals. Although the term cyberabuse is used throughout the document as an equivalent to cyberbullying for those over 18, it does not categorically capture the nature of harms associated with text-based abuse, such as online misogyny[6] or other forms of gender-based text-based abuse. Such exclusion leads to a limited view on what constitutes online harm (and, accordingly, 'online safety'). It also serves to reinforce their absence from the legal system – and from reporting & enforcement obligations. In turn, this results in a hierarchy of harms, where some harms such as IBSA are treated much more seriously than other forms of abuse.[7]

We acknowledge ad agree with the Government's position that 'online harms are complex and continue to evolve'.[8] Equally we do not advocate for a non-exhaustive approach – instead we support law reform based on categorisations of behaviours which are flexible but which recognize the harmful impact on wider demographics of internet users than reflected in the current document. Furthermore, a more inclusive approach would allow the capture of future online harms that may not be foreseeable at present, and this would go a long way in future-proofing legislation on online harms.

---

[3] These concerns are similar to the concerns surrounding the current UK Government proposals in the Online Harms reform agenda (12 February 2020): <https://www.gov.uk/government/consultations/online-harms-white-paper/public-feedback/online-harms-white-paper-initial-consultation-response>

[4] For example: An Act to Improve Enforcement of the Law in Social Networks (The Network Enforcement Act) 2017 (Germany); DCMS, 'Online Harms White Paper Initial Consultation Response' (12 February 2020) <https://www.gov.uk/government/consultations/online-harms-white-paper/public-feedback/online-harms-white-paper-initial-consultation-response> (UK); Harmful Digital Communications Act 2015 (New Zealand).

[5] Kim Barker & Olga Jurasz, *Online Misogyny as a Hate Crime: A Challenge for Legal Regulation?* (Routledge 2019) xiv.

[6] Amelia Lester, 'Ladylike: Julia Gilliard's Misogyny Speech' The New Yorker (9 October 2012) <https://www.newyorker.com/news/news-desk/ladylike-julia-gillards-misogyny-speech>; Sydney Morning Herald, 'Transcript of Julia Gilliard's Speech' (10 October 2012) <https://www.smh.com.au/politics/federal/transcript-of-julia-gillards-speech-20121010-27c36.html>.

[7] This disparity is particularly evident in – for example - the eSafety Commissioner's webpages on 'Online abuse targeting women' where the emphasis (at least in terms of space on the page) falls on image based abuses: <https://www.esafety.gov.au/women/online-abuse-targeting-women>

[8] Australian Government Department of Communications and the Arts, 'Online Safety Legislative Reform Discussion Paper (December 2019) 9. Hereafter, OSLR.

2. *Is the proposed statement of regulatory policy sufficiently broad to address online harms in Australia? Are there aspects of the proposed principles that should be modified or omitted, or are there other principles that should be considered?*

The expansion of existing obligations is a welcome one, especially the expansion of empowerment and upholding integrity. That said, the applicability starting point – whilst understandable – is concerning. The applicability point should be modified as a matter of urgency to include all platforms and sites where posting of content by users is permissible – especially to include e.g. newspaper comment pages / online discussion fora, and any platform where users are permitted (and encouraged) to share and post user-generated content.

3. *Is there merit in the BOSE concept?*

Yes but there is a risk that it becomes synonymous with only one aspect of the Online Safety Legislative Reform – particularly SbD. For the BOSE concept to have continued merit and to continue to be at the heart of the reform proposals, we suggest it should be subject to regular review by the relevant minister / oversight body (for example, every 24 months) given the evolution of both digital interactions, but also the associated and complex evolution of online harms. For full merit, the BOSE concept must also evolve. The assessment and re-evaluation of the BOSE concept should be closely linked to the transparency reporting mechanisms so that – for instance – if the transparency reporting highlights particular trends, these can be embedded into the BOSE concept, or the BOSE concept can be tweaked accordingly so ensure that the concept if introduced, remains relevant and workable.

4. *Are there matters (other than those canvassed in the Charter) that should be considered for the BOSE? Are there any matters in the Charter that should not be part of the BOSE?*

It is questionable to enshrine the voluntary benchmark for best practice in the BOSE framework particularly given the voluntary nature of this suggested by the Charter. If it is only 'basic' expectations, then the voluntary benchmark is likely futile.

5. *What factors should be considered by the eSafety Commissioner in determining particular entities that are required to adhere to transparency reporting requirements (e.g. size, number of Australian users, history of upheld complaints)?*

Transparency reporting requirements should be imposed on those platforms witnessing the greatest number of users, but also the greatest number of complaints – upheld or otherwise. We do not advocate for imposing these requirements on upheld complaints only – reporting complaints not upheld is as important as (if not more important) than those upheld. The transparency reporting requirements must feed in to both the review of the BOSE framework, but also the enforcement framework – otherwise, it is only a reporting mechanism for the sake of it, rather than for something of tangible value and input. Transparency reporting requirements would ideally be imposed on *all* entities allowing user-posted content, but this is likely to be unfeasible for smaller, more niche entities.

Drawing distinctions between entities to impose obligations on through user or subscriber numbers has been an approach previously utilised so it is not novel.[9] That said, we are sceptical of the benefits of distinguishing between platforms in this way – if there is a serious intent to address online harms, requirements should be imposed on all relevant platforms. We therefore are not supportive of the notion that some entities could be exempt from reporting requirements - equivalent requirements of a lesser frequency ought to be considered for smaller entities.

8. Is the proposed take-down period of 24 hours reasonable, or should this require take-down in a shorter period of time?

Generally speaking, the take-down period of 24 hours is reasonable – especially given its international benchmarking,[10] providing appropriate appeal and reinstatement mechanisms exist. A shorter take-down period is likely to be much less workable for all forms of harmful content that may pose a risk. Even where automated content moderation and review systems are in operation, they are not infallible and still currently require human oversight so as to not impinge upon free speech provisions.[11] Take-downs only operate in respect of the original content, and do not equate to stay-down, nor to reposted or repeated content elsewhere – these issues should also feed into the take-down system – and timescale. Input from platforms in respect of the technical aspects of take-downs is required, especially in light of recent announcements about the proactivity some platforms are purporting to take.[12]

11. Is the proposed application of the cyberbullying and cyber abuse schemes to designated internet services and hosting services, relevant electronic service and social media services, appropriate?

The proposed application of the cyberbullying and cyber abuse schemes should apply to all services hosting content – especially where such content is not automatically reviewed by the host.

12. *Is the proposed take-down period of 24 hours reasonable, or should this require take-down in a shorter period of time?*

This time period is reasonable to allow for a more comprehensive response from the providers – for instance, one that takes account of the contextual, socio-cultural and linguistic factors that may need to be considered before the content is taken-down. For instance, some content may not appear abusive or distressing to a person who lacks given socio-cultural or linguistic experience or knowledge but would amount to such for a person appreciating these differences. Therefore, a more nuanced approach is

[9] See for example Digital Economy Act 2010; Ofcom, 'Online Infringement of Copyright and the Digital Economy Act 2010, (28 May 2010) para 1.6. < https://www.ofcom.org.uk/consultations-and-statements/category-2/copyright-infringement>

[10] See, for example: European Commission, Commission Recommendation of 1.3.2018 on measures to effectively tackle illegal content online (C(2018) 1177 final).

[11] Sarah T Roberts, *Behind the Screen: Content Moderation in the Shadows of Social Media* (2019), 35.

[12] See for example, Faiza Patel and Laura Hecht-Felella, 'Evaluating Facebook's New Oversight Board for Content Moderation' Brennan Centre for Justice Briefing (27 November 2019) <https://www.brennancenter.org/our-work/analysis-opinion/evaluating-facebooks-new-oversight-board-content-moderation>.

needed to ensure that content that is taken-down actually conforms to these requirements and satisfies the given threshold for take-down.

Furthermore, whilst take-down may appear as a useful tool, it does not remedy the problem of online abuse – it merely makes it less visible to users. As such, measures other than legislation must also be taken to educate both young people and adults about being civil online and to enact gradual, long-term change in behaviours online.

13. Do the proposed elements of a definition of adult cyber abuse appropriately balance the protection from harms with the expectation that adults should be able to express views freely, including robust differences of opinion?

Freedom of expression cannot be an excuse for abusive or disruptive behaviour online. Online abuse can, and does, result in participatory harms which drive individuals away from freely expressing their views online. For instance, misogynistic abuse has had a silencing effect on many women and girls, effectively resulting in curtailing their equal rights to participate in public (here: online) sphere, as guaranteed under the UN Convention on Elimination of All Forms of Discrimination Against Women 1979.[13]

The question invites considerations of weighing the impact of online abuse vs freedom of expression in relation to adults only. This is a questionable positioning of the debate given that children and young people/ adolescents also have rights to be heard to express themselves freely.[14] Online abuse is likely to have silencing impact on young people too and is equally likely to affect their participatory rights.[15]

14. Should the penalties differ under a cyber abuse scheme for adults and the cyberbullying scheme for children?

No. It should also be taken into consideration that young people are not just victims of online abuse – very often they are the perpetrators and should be held accountable for their actions online.[16] There are limitations in promoting new cyber abuse framework for adults alongside the cyberbullying scheme for children if the penalties differ depending on age only. Online safety regulation proposals purports to make cyberabuse and cyberbullying equivalent within the new regulatory scheme – as such it is crucial that there are no disparities for what amounts to the same behaviour, but at a different age. The behaviours must be investigated and addressed in the same manner so as to prevent behaviours

[13] Australia ratified CEDAW on 28 July 1983.
[14] These are enshrined in the UN Convention on the Rights of the Child 1989. Australia ratified the CRC on 17 December 1990.
[15] Girlguiding, 'Girls' Attitudes Survey 2016' (Girlguiding, 2016) <https://www.girlguiding.org.uk/globalassets/docs-and-resources/research-and-campaigns/girls-attitudes-survey-2016.pdf>
[16] Martin Evans, 'Hundreds of children investigated for Twitter abuse and bullying each year new figures reveal' The Telegraph (29 May 2014) < https://www.telegraph.co.uk/news/uknews/crime/10861173/Hundreds-of-children-investigated-for-Twitter-abuse-and-bullying-each-year-new-figures-reveal.html>

at a young age continuing into adulthood, but also to prevent adult behaviours becoming accepted norms for younger generations.

19. Is the proposed application of the take-down powers in the revised online content scheme appropriate?

See above – answer to Q8.

Operating harmonised take-down powers in terms of responses, notices, and response times would be the easiest system for platform operators if the New Online Content Scheme is followed, and there are clear take-down procedures dependent on the content. Having different systems for different forms of content could become unduly burdensome on platforms and operators – particularly operators dealing with millions and billions of content posts each day. Given the volume of content likely to be captured by the online content scheme, keeping the powers as streamlined as possible is the ideal approach.

20. Are there other methods to manage access to harmful online content that should be considered in the new Online Safety Act?

Not that are tried and tested systems. The YouTube upload filter has shown some potential signs of progress but this presents alternative problems, and has also proved to be not infallible[17] – still requiring human oversight. The Facebook Oversight Board is also – purportedly – working on systems to deal with content moderation but will still be very heavily reliant on automated systems,[18] with further rigour occurring through the vastly increased number of moderators working globally to tackle content issues.[19] Fully automated systems miss the context of a particular post – it is therefore (currently) inadvisable to rely on fully automated systems.

22. Is the proposed take-down period of 24 hours for the online content scheme reasonable or should this require take-down in a shorter period of time?

See above – answer to Q8.

Given the harmful nature of online content, where take-down (and stay-down) is the preferred option for addressing it, it should fall wherever possible within a maximum take-down period of 24 hours. Class 2 content as outlined in the New Online Content Scheme should still be subject to take-down powers where the content is 'seriously harmful material' but the definitions of such material need to be carefully worded – and again, subject to regular review (either judicially, or governmentally).

[17] Louise Matsakis, 'YouTube doesn't know where it's own line is' WIRED (3 February 2018) <https://www.wired.com/story/youtube-content-moderation-inconsistent/>.
[18] Ibid Patel and Hecht-Felella at n12.
[19] Ibid Roberts at n11.

34. Is the requirement that 3rd parties be systemically and repeatedly facilitating the posting of cyberbullying or cyber abuse material, image-based abuse or hosting illegal or harmful content appropriate before the eSafety Commissioner can issue a notice to an ancillary service provider? Should a different threshold be contemplated?

Initially, the threshold seems adequate. Alongside the proposals outlined in the OSLR documentation, a graduated scheme of financial penalty is worthy of consideration in addition to delisting and deranking websites. These powers should be operated on a sliding scale dependent on severity and frequency of failures to tackle the posting / facilitation of posting such material. For instance, it may be that financial penalties are appropriate to impose (proportionate to the size of the operator) after the direct take-down powers have been utilised but where a problem persists. The delist and derank powers could then be utilised as a last-resort where there are systemic and well-documented persistent failings to act.

The threshold is rather vague at present – the requirement of 'systemic and repeated' is broad – what does this mean in practical terms? Does this mean that action will be taken where e.g. a percentage of reports suggest systemic failures? Or is it that this envisages a repeated failure where there is e.g. a maximum percentage of failures permissible? Or is the threshold something more directly quantifiable – would it, for instance, operate where there is a failure on 3 occasions to address the content? A different – fixed threshold in terms of *quantification* ought to be considered or implementation once there has been an identified problem with the platform or operator not addressing abusive, illegal or harmful content.

36. Are the eSafety Commissioner's functions still fit for purpose? Is anything missing?

The broad functions remain ambitious enough to suggest that the eSafety Commissioner still has a significant leadership role to play in spearheading campaigns addressing online, or, cyber abuses. That said, there is a clear set of hierarchical priorities evident in the functions and remit of the eSafety Commissioner, to the detriment of text-based abuses, and online misogynistic abuses. For instance, the eSafety Commissioners website highlights an overwhelming focus on IBSA (with specific but more minor mentions of fake accounts and cyberstalking but not to the same extent). This represents the manifestation of the difficulties in this area in that there is not an equal level of balance between image-based abuses online, and other, non-image-based abuses. This is further perpetuated by the Identification of some harms but not others. Such an approach creates a hierarchy of harms, and does not necessarily but women at the core of that hierarchy in contexts beyond image.

Ultimately the eSafety Commissioner – like any regulatory ambition in the broad sphere of cyberabuse is attempting to tackle something that is the equivalent of digital quicksand. Similarly, given the scale of online abuse and its vast parameters, and the volume of content posted daily, it is unlikely that the eSafety Commissioner will be realistically able to address all complaints, and every facet of this phenomenon.

39. What are the likely impacts, including resource implications, on other agencies and businesses of a new Online Safety Act?

The most significant likely impact here will fall on agencies, particularly NGOs that support victims of online abuse. Given that the majority of these groups will likely be charitable, the increased regulation and capture of online abuses, and intended increased reporting will manifest itself in greater funding needs to support victims – yet there is no mention of these support mechanisms beyond the legislative review within the proposals, and this is a significant omission.

More specifically, there are greater sources of funding support for certain categories of abuse than for others – notable in respect of funding for those suffering IBSA harms than other harms. Equally, where there are elements of technologically facilitated coercion and control in domestic violence scenarios, there is greater support available to such victims through the domestic violence support sectors than for other victims of e.g. text-based abuses where there is no domestic violence context. The disparities in the recognition of harms evident in the proposals, the eSafety Commissioners office, and the funding mechanism for supporting victims perpetuate the inequality of treatment of online harms, but also the inequality of support, and reinforce the harms inflicted.