FEEDBACK ON AN EXPOSURE DRAFT: ONLINE SAFETY BILL 2020

The Australian Government proposes to introduce a Bill for an Act relating to online safety for Australians. This includes the proposed "world-first cyber-abuse take down scheme for Australian Adults" (Australian Government 2020a, 1). It also includes a "blocking scheme" of material which is tantamount to, promotes, instructs, or incites abhorrent violent conduct (Exposure Draft, 95(1); Australian Government 2020b, 2). "Abhorrent violent conduct" is specified in Subdivision H of Division 474 of the *Criminal Code Act 1995* (Cth); it includes terrorism, murder or attempted murder, torture, rape, and kidnapping.

The core expectation of the proposed *Basic Online Safety Expectations* framework is that the provider enables end-user safety by taking "reasonable steps to minimise" cyber-bullying, cyber-abuse, non-consensual sharing of intimate images, Class 1 material, abhorrent violent material, or materials inciting promoting or instructing abhorrent violence, and to prevent children's access to Class 2 materials. It should do this through the provision of a reporting or complaints system. It should also report to the E-Safety Commissioner on various issues (Exposure Draft, 46(1)).

As a researcher investigating the problem of hate speech against women online (Australian Research Council project DE190100719), and after conducting six months of data collection (Ethics Ref No.: H0018328), I am keenly aware of the extensive varieties of abusive, hostile, and derogatory speech that individuals and groups may be subject to.

I welcome the aims of the proposed Bill. However, I would like to suggest some amendments as follows. Here, I draw your attention to the definition of "cyber-abuse" in Section 7(1). Subsection (b) stipulates that "an **ordinary reasonable person** would conclude that it is likely that the material **was intended** to have an effect of causing serious harm to a **particular** Australian adult." I have three concerns.

First, is the standard of an "ordinary reasonable person" the most appropriate standard for determining whether serious harm was (intended to be) occasioned? As has been persuasively argued by feminist theorists and critical race theorists, "reasonableness has no specific content," but can only be given meaning in a specific context (Dolkart 1994, 201). However, in western contexts "the dominant neutral principles of law… [reflect] white, male, heterosexual, middle-class norms" (Dolkart 1994, 175). Insofar as some people are targeted with cyber-abuse *because* they belong to an oppressed group (e.g. because they are women, or because they are people of colour), the "ordinary reasonable person" standard is at risk of failing to fully comprehend the severity of the harm that the abuse has caused. This is especially true for the evaluation of offensive material. If the severity of the harm is misinterpreted, then this may lead to the dismissal of complaints and perpetuation of said harm.

> Thus, I propose that subsection (b) be reworded as follows: "an ordinary reasonable person **in the position of [the target]** would conclude…", in line with the phrasing of subsection (c).

Second, I question the extent to which it is appropriate to include in subsection (b) the phrase "the material **was intended** to have an effect…". If it is possible to demonstrate that the material in question has in fact caused a substantial harm, why is this not sufficient for determining an instance of cyber-abuse? I see no reason for *accidentally harmful material* to be classed outside the scope of cyber-abuse.

> Thus, I propose that subsection (b) be reworded as follows: "the material **had an effect**…".

This is consonant with a victim-centred approach to understanding 'harm' (Powell & Henry 2017).

Third, I would like to highlight some limitations with the provision that the material has "an effect of causing serious harm to a **particular** Australian adult." The offending material must be menacing, harassing, or offensive leading to serious harm, where "serious harm" is understood, under Section 5, to be either physical harm or harm to a person's mental health, wherein that may be serious psychological harm or serious distress.

While it is undoubtedly true that individuals who are cyber-abused can experience extraordinary distress, and that cyber-abuse can be used to target people indiscriminately, it is also true that certain collectives—again, women, people of colour, and others—face a heightened risk of encountering hostilities targeting them *as members of that social group* due to the ongoing web of discriminatory practices and prejudicial attitudes they have endured in the past, and which continue to affect them in the present. Material of this nature—expressions which systemically discriminate on the basis of group membership—is known as "hate speech" (Gelber 2019). Note: for the purposes of the below discussion I am referring to expression which meets these criteria, but which does not meet the threshold constituting 'abhorrent violent material'.

Hate speech and cyber-abuse do not appear to be synonymous, though it is clear that some expressions of hate could count as cyber-abuse were they to be of the specified kind (menacing, harassing, offensive) and to have the causal effect of serious harm on an identified Australian individual. Even so, this fails to account for two things: (a) it does not protect groups from harm; and (b) the conception of harm is too narrow.

I urge the Government to consider that the harms so-described here are not the only serious harms that can result from menacing, harassing, and offensive materials being hosted and shared online.

(a) In addition to serious harms *to the individual*, it is also possible for a *group* to be harmed by online material. When hate speech is circulated, and when that material is permitted to remain in public places (including public places in cyberspace, like social media), this sends a message to all users that certain groups may be (or deserve to be) the subject of derision, denigration, and animosity. It also sends a message to all users that certain other groups are "normal", or even "better" or "best". This harms the dignity status of the denigrated group, and the protection of this status is best served by the removal of hateful expressions (Waldron 2012).

(b) When harassing, menacing, and offensive material takes *groups* as its subject, it is not necessarily the case that specific individuals are harmed in the ways specified in the Exposure Draft. As mentioned above, (a) harm to groups is primarily a dignitarian harm that maintains unjust social hierarchies in contexts of already-existing inequality. However, when such group-oriented material proliferates in online social spaces (e.g. Facebook comments responding to the publication of a news article, as I have studied) it can create a *hostile environment* for members of the target group, and the constitution of a hostile environment for members of oppressed groups ought itself to be seen as a form of unjust discrimination, i.e. *a serious harm* of another kind (McGowan 2020). A hostile environment can stop people from participating in public debate, from visiting particular platforms, and even from having a "public presence" online altogether.

Since the Australian Government has few legislative tools to deal with hate speech against women (D'Souza et. al. 2018; de Silva 2020), it behoves the Government to either: i) amend its definition of serious harm to include group harms; and/or ii) add further provisions to include "hate speech" as cyber-abuse of groups; and/or iii) introduce additional legislation protecting women and other vulnerable collectives from "hate speech" equivalent to the civil provisions against racial hatred in the *Racial Discrimination Act 1975* (Cth).

It is immensely encouraging to see that the Government is taking proactive steps to ensure the safety of Australians online; it is important that Australians are protected from serious harms in this space, and that the legislation does not leave any member of our community more vulnerable than another.

Dr Louise Richardson-Self.

REFERENCES CITED:

Australian Government. 2020a. "Fact Sheet — Online Safety Bill" *Consultation on a Bill for a new Online Safety Act* <https://www.communications.gov.au/have-your-say/consultation-bill-new-online-safety-act> (accessed 04/02/21).

Australian Government. 2020b. "Online Safety Bill—Reading Guide." *Consultation on a Bill for a new Online Safety Act* <https://www.communications.gov.au/have-your-say/consultation-bill-new-online-safety-act> (accessed 04/02/21).

D'Souza, Tanya; Griffin, Laura; Shackleton, Nicole; and Walt, Danielle. 2018. "Harming Women With Words: The Failure of Australian Law to Prohibit Gendered Hate Speech." *UNSW Law Journal*. 41 (3): 939–976.

de Silva, Anjalee. 2020. "Addressing the Vilification of Women: A Functional Theory of Harm and Implications for Law." *Melbourne University Law Review* 43 (3): 1–46.

Dolkart, Jane. 1994. "Hostile Environment Harassment: Equality, Objectivity, and the Shaping of Legal Standards." *Emory Law Journal*. 43: 151–244.

Gelber, Katharine. 2019. "Differentiating Hate Speech: A Systemic Discrimination Approach." *Critical Review of International Social and Political Philosophy*. DOI:10.1080/13698230.2019.1576006.

McGowan, Mary Kate. 2020. *Just Words: On Speech and Hidden Harm*. Oxford: Oxford University Press.

Powell, Anastasia; and Henry, Nicola. 2017. *Sexual Violence in a Digital Age*. London: Palgrave Macmillan.

Waldron, Jeremy. 2012. *The Harm In Hate Speech*. Cambridge: Harvard University Press.

LEGISLATION:

Criminal Code Act 1995 (Cth).

Online Safety Bill 2020 [Exposure Draft]

Racial Discrimination Act 1975 (Cth).