SUBMISSION ON THE BILL FOR A NEW ONLINE SAFETY ACT

Minderoo Tech & Policy Lab, UWA Law School

Contents

1 Introduction

This submission responds to the exposure draft of the *Online Safety Bill 2020* published by the Department of Infrastructure, Transport, Regional Development and Communications, having particular regard to the 2019 *Online Safety Legislation Reform – Discussion Paper* that informed the draft Bill.

The overarching aim of the proposed legislation is to improve safety online. Achieving that aim will require building consensus, not only with the tech platforms who operate online communities, but with the citizens who are members of those online communities. As presently formulated, the proposed legislation fails to strike an effective balance between those needs. That is, it prioritises the needs of the companies operating online platforms and does not sufficiently empower the communities who use those platforms. The risks associated with this are twofold. First, by leaving tech companies so much leeway in establishing their own accountability standards, the present regime risks a regime for the protection of online safety which will not be effective. Second, by failing to provide robust democratic mechanisms for community participation in setting those standards, we risk a lack of community acceptance of – and compliance with – those standards.

The efforts of the eSafety Commissioner are laudable. However, no single body will be able to ensure online safety. Achieving this outcome requires building community consensus about standards and behaviour online. Indeed, as the eSafety Commissioner's 2020 report on *Online Hate Speech* found,[1] the vast majority of Australians surveyed agreed with the propositions that 'We will need to do more than introduce additional legislation to prevent the spread of hateful content online' and that 'Everyone has a role in tackling hateful content'.

The message here is clear: Australians want more action and believe they have a role to play in taking that action. This legislation is far more likely to achieve its purposes if it successfully facilitates community engagement. Recognising and responding to this need is a core objective of this submission.

In broad terms, the submission contends that the proposed legislative regime might be improved in four main ways: (1) implementing mechanisms to ensure that the Basic Online Safety Expectations (BOSE) contain standards which are set democratically by the communities whose expectations they seek to protect; (2) mandating compliance with the BOSE; (3) ensuring that the highest safety standards with respect to privacy apply by default, regardless of the age of the user; and (4) future-proofing the Bill through regulating online misinformation harms, as in contemporaneous reforms in the UK's proposed *Online Safety Bill*.

As well as the broader issues above, this submission also considers some of the detail of the draft exposure Bill. It identifies two interpretive challenges which arise from the present drafting of that Bill. First, the subjective/objective test used in the s7 definition of cyber-abuse material. Second, the definition of intimate image in s15(2). The submission will recommend redrafting s7 to

[1] eSafety Commissioner, 'Online Hate Speech: Findings from Australia, New Zealand and Europe' (Report, eSafety Commissioner, 2020) <https://www.esafety.gov.au/sites/default/files/2020-01/Hate%5C%20speech-Report.pdf>.

avoid difficulties with the *mens rea* component of that test. It will also suggest amending s15(2) to simplify and clarify the application of that section to persons who are transgender or intersex persons not identifying as female.

To that end, this submission makes six key recommendations set out below.

2 Recommendations

1. The legislation should ensure that the Basic Online Safety Expectations (BOSE) are set democratically, by the communities whose expectations they seek to protect. Such an approach should:
   (a) Set expectations which are targeted at particular classes of online services;
   (b) Involve community users of those services in the formulation of the BOSE; and
   (c) Ensure that the mechanisms by which the BOSE are set involve ongoing engagement and re-review of the standards.
2. The proposed legislation should be amended to ensure that the BOSE are legally enforceable standards in the first instance.
3. The proposed legislation should be amended to require that the protection of the highest safety standards with respect to privacy apply to all users by default.
4. The proposed legislation should consider adopting the protections aimed at curbing online misinformation which are being developed in other jurisdictions, such as the UK's *Online Safety Bill*. Such a mechanism should be informed by the same principles of community engagement in formulation, implementation, and review recommended in (1) above.
5. Section 7(b) of the proposed legislation be amended either to remove the test of intention and provide for a strict liability offence, or to clarify that 'all the circumstances' related to the material are relevant to interrogation as to the likely intention.
6. Section 15(2) be redrafted to simplify and clarify the application of that section.

3 Who We Are – Minderoo Tech & Policy Lab

The Minderoo Tech & Policy Lab is a research institute headquartered at The University of Western Australia. The Lab is directed by legal scholar Associate Professor Julia Powles and technologist Associate Professor Jacqueline Alderson, who lead an interdisciplinary team of researchers that specialise in the

development and regulation of emerging technologies. This submission was led by research fellow Tomas Fitzgerald.

The Lab commenced operations in September 2020 as a core node in an international tech impact network focused on tackling lawlessness in the technology ecosystem, with partners including the University of Cambridge, the University of California Los Angeles, New York University, the University of Oxford, the Australian National University, the University of Sydney, and more.

The Lab pursues twin objectives: promoting and protecting rights for individuals and communities faced with the harmful consequences of digital technologies and data-informed systems; and providing a robust pro-innovation environment and use-cases for the stimulation of civic tech development in the public interest.

The Lab acknowledges the support of Australian charity Minderoo Foundation in the creation of the Lab. We maintain the highest standards of academic integrity and are committed to the autonomy and independence of our researchers to pursue work free of external influence.

4 The Basic Online Safety Expectations and Industry Codes

One of the key mechanisms in the proposed Bill is the Basic Online Safety Expectations (BOSE), which are determined by legislative instrument. As the draft Bill provides:

> **45 Basic online safety expectations**
>
> *Social media service*
>
> (1) The Minister may, by legislative instrument, determine that the basic online safety expectations for a social media service are the expectations specified in the determination.
>
> *Relevant electronic service*
>
> (2) The Minister may, by legislative instrument, determine that the basic online safety expectations for each relevant electronic service included in a class of relevant electronic services specified in the determination are the expectations specified in the determination.
>
> *Designated internet service*
>
> (3) The Minister may, by legislative instrument, determine that the basic online safety expectations for each designated internet service included in a class of designated internet services specified in the determination are the expectations specified in the determination.

Determination does not impose a legally enforceable duty

(4) A determination under this section does not impose a duty that is enforceable by proceedings in a court.

The goal of these instruments is to '[set] out Government's expectations of industry' and provide a 'Point-in-time, voluntary benchmark for best practice'. However, as the Discussion Paper notes, '[t]he Government is not proposing to impose sanctions for non-compliance with the proposed basic online safety expectations at this stage, though reserves the right to explore this option in future if expectations are not being met'.

While there are reporting requirements as against the proposed BOSE, subsection (4) explicitly clarifies that the BOSE will not impose legally enforceable duties, aside from the duty to report against those criteria. The consequence of this is that the BOSE will remain aspirational, and the only enforceable principles will be those set by industry themselves. That is, the Industry Codes – discussed below – will be the mechanism which sets out the standards which industry can be held to by the enforcement provisions in the proposed Bill.

The Discussion Paper asks:

3. Is there merit in the BOSE concept?
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?

To the extent that the BOSE attempts to set minimum standards across industry by legislative instrument, there is merit in the concept. However, as presently proposed the BOSE suffers two serious flaws. First, it does not go far enough in terms of capturing the community's expectations of online service providers. Second, the lack of enforcement mechanism renders the BOSE purely aspirational. What merit there is in the BOSE concept is likely to be eroded if the mechanism for implementing it remains as set out in the draft bill. Rather, the BOSE ought to be a more robust statement of the community's expectations and should be enforceable at the outset.

The Discussion Paper identifies that the general trend towards voluntary or industry-led compliance mechanisms reflects two concerns: a desire to minimise the regulatory burden on industry; as well as the political sensitivity of imposing stricter requirements, including on questions such as freedom of speech. As regards the former concern it is worth recalling Briggs' position that, 'I am drawn to say that the view sometimes promulgated by the online industry that increased regulation will damage innovation, is complete bunkum.'[2]

[2]Lynelle Briggs, 'Report of the Statutory Review of the Enhancing Online Safety Act 2015 and the Review of Schedules 5 and 7 to the Broadcasting Services Act 1992 (Online Content Scheme)' (Report, Department of Infrastructure, Transport, Regional Development and Communications, 2018) 40 <https://www.communications.gov.au/publications/report-statutory-review-enhancing-online-safety-act-2015-and-review-schedules-5-and-7-broadcasting>.

As regards the question of the appropriate limits of regulatory intervention on online content, and hence intervention in politically sensitive areas, the relevant question is not whether intervention *per se* is desirable. Australia does not have, and has never had, the values of the American polity which privilege the right to free speech – no matter if offensive, untrue, harmful, racist, or anti-democratic – over all other considerations. Nor has Australia had a particularly censorious history. Regulation of speech in Australia is rather a question of proportionate response to harms. Hence robust laws against, among other things, defamation, hate speech, misleading and deceptive conduct, as well as broad mechanisms for regulating film and literature classification and even news media publishing.

Despite the existence of robust laws around classification of film, literature and video games, and news media publishing, there is no general sense that Australia is engaged in censorship, relevantly speaking. This is to be distinguished from experiences elsewhere in the world, where similar regulations are used by States to censor dissent. The distinction between objectionable censorship and appropriate regulation is a question both of degree, but more importantly of process.

The Australian Classification Board represents an excellent example of a robust, democratic mechanism for ensuring that the community's expectations about media classification are given effect by regulation. Decisions about classification are made by the independent Classification Board. Section 48(2) of the *Classification (Publications, Films and Computer Games) Act 1995* (Cth) requires that:

> (2) In appointing members, regard is to be had to the desirability of ensuring that the membership of the Board is broadly representative of the Australian community.

Further, the National Classification Code, which is applied by the board, requires balancing a set of principles which explicitly centre community expectations:

> 1. Classification decisions are to give effect, as far as possible, to the following principles:
>
> (a) adults should be able to read, hear, see and play what they want;
>
> (b) minors should be protected from material likely to harm or disturb them;
>
> (c) everyone should be protected from exposure to unsolicited material that they find offensive;
>
> (d) the need to take account of community concerns about:
>
> (i) depictions that condone or incite violence, particularly sexual violence; and
>
> (ii) the portrayal of persons in a demeaning manner.

The strength of Australia's classification system is that it implements an independent board of classifiers, requires that the board be broadly representative of the community, and requires classification assessments against a code which centres the community's expectations. These features have meant that the operation of the classification regime has enjoyed strong community support. An exercise which might – in theory – be highly politically contentious is, in practice, a mundane exercise which, crucially, enjoys broad support within the community.

It is this last feature of the operation of the classification regime which regulators seeking to implement an online safety regime ought to be most interested in emulating. As the eSafety Commissioner noted when reporting on online hate speech, 70% of Australians agree with the statement that 'everyone has a role in tackling hateful content'[3] That report also found that 'the overwhelming majority of people support action to check the spread of online hate speech including the introduction of legislation and getting social media companies to do more.'[4] The existing classification board provides an excellent model for achieving robust community consensus on an otherwise potentially contentious political issue. This model could be replicated for formulating the BOSE. The draft Bill could be expanded to create a body, like the classification board, drawn from community members given fixed term appointments tasked with formulating and reviewing the content of the BOSE. Such a board, tasked with giving effect to the community's strong sentiment that regulation proportionate to the harm, is both necessary and desirable. Like the classification board, this would permit essentially political decisions to be made in a non-partisan manner focused primarily on giving effect to the community's expectations. It would be an effective and robust democratic response to the harms occasioned by these technologies.

Such a board would also act as an essential democratic safeguard, giving regulators the confidence to back the BOSE with mechanisms which permit direct legal enforcement of its requirements. This mechanism would permit regulators to meet the community's clear desire for robust regulation while avoiding the danger that regulation in this area might become a partisan political tool.

Implementing a greatly strengthened BOSE backed by a robust, democratic mechanism for setting and reviewing its content and permitting direct legal enforcement of that content would go a long way to meeting the conceptual and practical challenges of regulating online safety. It is also likely – as the classification regime has – to build broader consensus within the community about the appropriate standards for online communication.

4.1 Lessons From Other Industry Codes of Conduct

The alternative to implementing a stronger BOSE, backed by the capacity for direct legal enforcement is the creation of a second tier of Codes of Conduct.

[3] eSafety Commissioner, above n 1, 7 <https://www.esafety.gov.au/sites/default/files/2020-01/Hate%5C%20speech-Report.pdf>.

[4] Ibid 6.

This is the regulatory mechanism set out in the proposed Bill. Given that the stated goal of the proposed legislation is to 'spark the creation of new industry codes to address harmful online content',[5] it is worth considering: (1) whether industry co-design is likely to be appropriate in this case; and (2) whether industry co-design of regulation has worked in similar areas.

4.1.1 Is Industry-Led Design Desirable?

As the Briggs report reviewing Australia's existing online safety regime observed, industry's commitment to online safety waxes and wanes with time. It advised that '[t]he lesson to be drawn from this is that the level of industry commitment to online safety is fragile and unreliable, and needs to be shored up by being given a legislative basis.'[6] It is worth reproducing Briggs' characterisation of the issue:

> By and large, declining rates of public trust and rising levels of outrage are strong indicators that the Australian community's hopes have been shattered in terms of their belief that people, industry and businesses will exhibit conduct at a level commensurate with community expectations. The dilemma facing society is that people no longer feel that they can rely on industry, business or even the church to do the right thing, let alone individuals working within the system.
>
> The social license to operate of some business sectors, many companies and institutions is being increasingly challenged as the community demands higher standards which better reflect their expectations of good behaviour and appropriate practice. It should therefore come as no surprise to the online industry that the tide of change is against it–with the community calling on the Government to provide higher levels of intervention to control and penalise misconduct online, whether it be malfeasance or neglect, and to protect Australians more generally.[7]

It is against this backdrop that we must judge the fundamental proposal of the *Online Safety Bill 2020* that the design of the regulatory framework be industry-led. In that regard, the BOSE are to be contrasted with the Industry Codes provided for by Part 9 Division 7 of the proposed Bill. As the Bill notes:

> **137 Statement of regulatory policy**

[5] Department of Communication and the Arts, 'Online Safety Legislation Reform Discussion Paper' (Discussion Paper, Department of Communication and the Arts, December 2019) 1 <https://www.communications.gov.au/file/48929/>.

[6] Briggs, above n 2, 38 <https://www.communications.gov.au/publications/report-statutory-review-enhancing-online-safety-act-2015-and-review-schedules-5-and-7-broadcasting>.

[7] Ibid.

> (1) The Parliament intends that bodies or associations that the Commissioner is satisfied represent sections of the online industry should develop codes (industry codes) that are to apply to participants in the respective sections of the industry in relation to their online activities.

Those Industry Codes then become the basis for remedies available under the proposed Bill, specifically ss144 and 145. These sections will give direct legal effect to the Industry Codes – in practice making those the mandatory standards which can be readily enforced by judicial intervention. By contrast, breaches of the BOSE attract no parallel remedies, and are relevant only as regards the various reporting requirements set out in Part 4 Division 3 of the proposed Bill.

In essence, s137(1) clarifies that the proposed regime of Industry Codes are to be industry-led. To be sure, the Commissioner must be satisfied that those codes were developed in a manner which invited public submissions and considered those submissions by virtue of ss140(1)(e)(i) and (ii). However, that is the extent of any requirement for consultation. The relevant question for policymakers is this: do we expect that industry-led codes are likely to be sufficient to address the public concerns articulated by Briggs? Concluding in the affirmative would seem to be a triumph of hope over experience.

Further, unlike the BOSE and despite the requirements to consider public comment contained in ss140(1)(e)(i) and (ii), industry led codes suffer from an intrinsic democratic deficit. By their nature, they place development of standards in the hands of industry, not the community. Experience has demonstrated that industry led standards have not been effective. One of the main reasons they have not been effective is that they do not encourage broad community consensus or buy-in about the issues they seek to address. They have been demonstrably unsuccessful in building a broader culture of safety online. One of the main virtues of the approach proposed by this submission is that it argues for the implementation of a democratic, community-centred approach to formulating the BOSE that is much more likely to build that consensus, and hence to change community behaviour, in practice.

4.1.2 Have Other Industry Codes of Conduct Worked?

In addition to the lack of success achieved by industry's existing codes of conduct in relation to online safety, experience in similar areas demonstrates the limits of industry-led voluntary codes of conduct. One useful illustration was the Voluntary Industry Code of Conduct on Body Image. That Code of Conduct, 'was developed by the National Advisory Group on Body Image and provides a list of best practice principles to guide professionals in the media, advertising and fashion industries about body image.'[8] That code was established in order

[8] *Minister calls for industry action on body image* , (27 June 2010) Department of Education, Skills Employment <https://ministers.dese.gov.au/ellis/minister-calls-industry-action-body-image>.

to address the harms which arose from the ubiquity of 'photoshopped' images, particularly the 'body image pressures on young people in particular'.

Despite those laudable intentions, the voluntary code simply did not work. As de Freitas et al note:

> Unfortunately, these findings appear to confirm that voluntary efforts such as the Australian Voluntary Industry Code of Conduct on Body Image have had a placatory and short-lived impact on media practices, at least in the print media. Initiatives by government, industry and other sectors are to be commended, yet there is a clear need to ensure these initiatives are effectively implemented and evaluated in an ongoing manner to ensure sustained impact. For example the Australian Voluntary Industry Code of Conduct on Body Image initiated in 2010 was an admirable achievement. However, its voluntary nature limited implementation, and it was essentially abandoned upon a subsequent change in government in 2013.[9]

Mia Freedman, the former chair of the National Body Image Advisory Group who first recommended implementation of the voluntary code puts the issue more starkly:

> That's why I've changed my mind. Voluntary doesn't work. The disclosure of digitally altered images must be mandatory. It's just not going to happen otherwise.[10]

The parallels between the issues are plain. Both are attempts to regulate harm caused by new technology. Both technologies are ubiquitous and widely adopted. Both have a poor history of industry self-regulation and were faced with increasing public demands for further regulatory intervention. The question is: would we expect the experience of regulating against harms from social media to be relevantly different from regulating harms associated with image manipulation?

It is true that the proposed Bill does set out a mechanism for enforcing the content of industry-led codes in ss144 and 145. However, as written the extent of enforcement for the BOSE is a reporting requirement. It follows that the only genuine mechanism for enforcing the regulation in any particular instance of a breach is by reference to the industry-led codes. There is, at present, no genuine mechanism for ensuring that those codes contain robust and democratically determined statements of the community's expectations. Without such a mechanism, it is likely that industry's response will – as it has in the past – continue to fall short.

Consequently we make the following recommendations:

[9] Catarina de Freitas, Helen Jordan and Elizabeth K. Hughes, 'Body image diversity in the media: A content analysis of women's fashion magazines' (2018) 29 *Health Promotion Journal of Australia* 251–256 254.

[10] Ibid.

1. The legislation should ensure that the Basic Online Safety Expectations (BOSE) are set democratically, by the communities whose expectations they seek to protect. Such an approach should:
   (a) Set expectations which are targeted at particular classes of online services;
   (b) Involve community users of those services in the formulation of the BOSE; and
   (c) Ensure that the mechanisms by which the BOSE are set involve ongoing engagement and re-review of the standards.
2. The proposed legislation should be amended to ensure that the BOSE are legally enforceable standards in the first instance.

5 Privacy Concerns

There is presently significant law reform being undertaken in the area of privacy. In particular, the Commonwealth Attorney-General's department is conducting a review of the *Privacy Act 1988* (Cth). We have made a detailed submission to that review.[11]

Many of the issues that arise in relation to privacy reform are relevant to the question of online safety. Hence, it would be desirable if the portions of the *Online Safety Bill* that relate to privacy were congruent with that regime.

There are nevertheless some specific advances in the context of online safety that need not await the ultimate position taken in the *Privacy Act*. In particular, the *Online Safety Discussion Paper* canvassed mandating that products marketed to children default to the highest level of privacy settings:

> The Government is looking for industry to ensure that products marketed to children default to the highest level of privacy and safety at the outset, and to enable consumers to set and adjust these controls as they wish. It would be preferable to have these enhanced safety features developed and implemented voluntarily through an industry wide commitment to safety, consistent with the SbD [Safety by Design] principles and basic online safety expectations. However, in the event that a sector of the industry or particular service providers don't adopt this as a standard practice, the Government will consider the merits of empowering the eSafety Commissioner to specify, by legislative instrument, that particular types of service, or individual service providers with services marketed to children, default to the most restrictive privacy and safety settings.

[11] Fitzgerald et al., 'Submission to the Review of the Privacy Act 1988 (Cth) Issues Paper' (Submission,) <https://www.ag.gov.au/sites/default/files/2020-01/minderoo-tech-and-policy-lab-university-of-western-australia-law-school.PDF>.

In our view it is desirable for the eSafety Commissioner to be empowered with this capacity at the outset. Indeed, we recommend going further still. All online products ought to default to the highest level of privacy and safety settings, regardless of the age of their target market. Such a move would reflect community expectations and be consistent with international trends, such as reforms to align with the world's leading privacy instrument, the European Union's General Data Protection Regulation (GDPR).

Australians' attitudes to privacy online were surveyed in the eSafety Commissioner's September 2020 report *Building Australian Adults' Confidence and Resilience Online*. That report found that 'Seven in ten Australians ranked privacy as *the top issue for technology companies to address*, saying that it's important they ensure *the highest privacy settings are in place by default* (73%)'. Consequently, the starting point for discussion must be the fact that there is overwhelming community support for requirements that tech companies ensure that their products default to the highest privacy and safety standards.

As well as this, there are good pragmatic reasons for insisting that the highest default privacy protections apply across all age groups, not only children. Requiring differential treatment of consumers based on their age would in practice require tech companies to either ascertain the age of each of their users or prospective users, or to treat certain companies or products as marketed at youth in general. In the first case, this requires collection of information which may itself become a risk to privacy. This risks a perverse outcome; increasing the risk of privacy breaches so as to permit tech companies to more easily discern whom additional privacy protections must be extended to. In the second case, there are frequently large crossovers in the use of online services by particular age groups – particularly social media platforms. Any such categorical distinction would in many cases be arbitrary.

Further, it is technically less cumbersome to apply blanket, rather than targeted, privacy protection defaults. Hence it is no more onerous to require that all users be extended the protection of the highest level of default privacy protection than to require any one category of users to be so protected.

International regulatory regimes – such as the GDPR – already require in practice that online services default to the highest level of privacy protection:

> **Article 25 Data protection by design and by default**
> 2. The controller shall implement appropriate technical and organisational measures for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed. That obligation applies to the amount of personal data collected, the extent of their processing, the period of their storage and their accessibility. In particular, such measures shall ensure that by default personal data are not made accessible without the individual's intervention to an indefinite number of natural persons.[12]

[12]European Parliament, *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC*

Consequently, requiring that services default to the most secure privacy setting will harmonise Australia's regulatory regime with that of the European Union. It is also consistent with the Safety by Design ('SbD') principles which the eSafety Commissioner seeks to give effect to:

> **Principle 2-1** Provide technical measures and tools that adequately allow users to manage their own safety, and that are set to the most secure privacy and safety levels by default.[13]

Finally, it should be noted that requirements for certain *default settings* are clearly distinct from mandating *restrictions* on possible privacy settings. In other areas of regulation, such as the new Online Content Scheme,[14] the proposed legislation will effectively allow the prohibition of certain content, based on balancing competing claims about the harms associated with that content against claims as to legitimate reasons for permitting it.[15] In this case, no such question of balancing arises, as no restrictions or prohibitions are being proposed. Individuals – adults in particular – will retain the capacity to opt out of default privacy settings if they so choose. Against this backdrop the argument for mandating particular default settings is all the more compelling.

Taken together, these considerations speak in favour of extending the requirement for the highest level of default privacy protection beyond children and to all users.

For these reasons we make recommendation three:

3. The proposed legislation should be amended to require that the protection of the highest safety standards with respect to privacy apply to all users by default.

(General Data Protection Regulation) (4 May 2016) <http://data.europa.eu/eli/reg/2016/679/oj/eng>.

[13] eSafety Commissioner, *Safety by Design* (2021) <https://www.esafety.gov.au/about-us/safety-by-design>.

[14] Department of Communication and the Arts, above n 5, 39 <https://www.communications.gov.au/file/48929/>.

[15] In relation to the Online Content Scheme, we are aware that some groups are organising submissions to this inquiry which raise issues as to the appropriateness of that scheme. In this regard, we note that there is a concurrent review of the National Classification Scheme. As such we propose to not address the question of the content of that scheme. However, it is worth observing that whatever the content of that scheme, is is desirable that any regulation which purports to permit the restriction of that content is well targeted and capable of being implemented. As a matter of policy, it is plainly undesirable that laws be enacted which are simply impossible to enforce. Such laws undermine the public's confidence in the legal system more generally and are apt to create opportunities for arbitrariness, and even corruption among bodies tasked with their notional enforcement. As influential legal theorists Lasswell and McDougal observed in their seminal text *Jurisprudence for a Free Society*, law is best understood as the union of authority and control. Just as control without authority is objectionable as authoritarianism, so too is the exercise of authority without control objectionable as mere pretence.

6 Online Safety and Misinformation

The Discussion Paper notes that 'A number of overseas jurisdictions have also taken proactive measures to improve the accountability of digital platforms for content and behaviour on their services'. Key among these are those nations who have undertaken, or are proposing to undertake, regulation of the harms of online misinformation. The UK Government's response to the *Online Harms White Paper* provides the most salient example:

> This response to the Online Harms White Paper sets out plans for a new duty of care to make companies take responsibility for the safety of their users. It builds on our manifesto commitment to introduce legislation to make the UK the safest place in the world to be online but at the same time defend freedom of expression.
>
> The legislation will define what harmful content will be in scope. Principally, this legislation will tackle illegal activity taking place online and prevent children from being exposed to inappropriate material. But the legislation will also address other types of harm that spread online – from dangerous misinformation spreading lies about vaccines to destructive pro-anorexia content.[16]

The UK is not alone, with various proposals for regulation of online misinformation globally, including from Canada, the EU, France, Germany, Singapore, and Brazil. However, despite this burgeoning global trend, the proposed *Online Safety Bill* does not extend to the regulation of harmful misinformation.

It is acknowledged that regulatory intervention in this space is recent and evolving, and has in many ways superseded the context of Australia's review of its online safety regime. However, this is an area in which threats evolve rapidly, and regulatory response must be rapid to be effective. Formulating provisions to regulate the harms associated with online misinformation would be an effective step to future-proof Australia's regulatory system. Moreover, such a response could be coupled with the democratic mechanisms set out above to give Australians a direct say in what harms are regulated and the manner in which they are regulated. By involving the community in the design, implementation, and ongoing monitoring of the systems for regulating harms online, the regulations would be implementing a democratic safeguard which is likely to obviate many of the difficulties associated with regulation in this area.

Hence, we make recommendation four:

[16] *Online Harms White Paper: Full government response to the consultation* , (2020) GOV.UK <https://www.gov.uk/government/consultations/online-harms-white-paper/outcome/online-harms-white-paper-full-government-response>.

4. The proposed legislation should consider adopting the protections aimed at curbing online misinformation which are being developed in other jurisdictions, such as the UK's *Online Safety Bill*. Such a mechanism should be informed by the same principles of community engagement in formulation, implementation, and review recommended in (1) above.

7 Difficulties with Current Definitions of Cyber-Abuse Material and Image Based Abuse

The Discussion Paper asks:

> 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?

Broadly speaking the proposed elements do strike an appropriate balance. However, the proposed elements do give rise to a potential area of confusion which we suggest ought to be addressed. Specifically, the test for what effectively amounts to the *mens rea* component of the civil offence might be constructed in a way which excludes ostensibly relevant evidence from consideration. We propose amending that test to avoid this difficulty.

7.1 The Definition of Cyber-Abuse Material

The exposure draft proposes a new section 7:

> **7 Cyber-abuse material targeted at an Australian adult**
>
> (1) For the purposes of this Act, if material satisfies the following conditions:
>
> (a) the material is provided on:
>
> (i) a social media service; or
>
> (ii) a relevant electronic service; or
>
> (iii) a designated internet service;
>
> (b) 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;
>
> (c) an ordinary reasonable person in the position of the Australian adult would regard the material as being, in all the circumstances, menacing, harassing or offensive;

(d) such other conditions (if any) as are set out in the legislative rules;
then:

(e) the material is cyber-abuse material targeted at the Australian adult; and

(f) the Australian adult is the target of the material.

Subsection s7(1)(b) is the provision of most concern. That section requires that it be established that a person who is alleged to have committed the civil offence acted in such a manner that the 'ordinary reasonable person' would take a particular view of the intention behind the material. Specifically, that the intention was 'to have an effect of causing serious harm to a particular Australian adult'. Importantly, this is not a subjective test of the alleged offender's *actual* intention. Rather it is an objective test. It asks whether a third party – the ordinary reasonable person – would believe it 'likely that the material was intended to have' such an effect.

It is not clear on the face of s7(1)(b) what evidence will be relevant when it comes to establish that intention. The language of the test is drawn from the law of defamation, where the ordinary reasonable person test is used to establish 'whether a published matter is capable of being defamatory'.[17] In that context the test explicitly is not an examination of what the 'words or images in fact say or depict'; instead it asks 'what a jury could reasonably think they convey to the ordinary reasonable person...'[18]

Two difficulties arise with the application of this test. The first is the general difficulty with the legal fiction of the 'ordinary reasonable person' articulated by Kirby J in his assenting judgment in *Favell v Queensland Newspapers Pty Ltd*. His Honour's observation that 'generally speaking, the law is moving away from fictions and in the direction of substance and reality'[19] is of particular relevance to the drafting of new legislation. Since the operation of a legal fiction is always to have some practical effect on the manner in which a case is plead or proved, it is preferable to address those concerns directly, rather than indirectly.

The force of this more theoretical difficulty is best understood in the context of the second, more practical, difficulty that this test invites. Subsection 7(1)(b) seems to be aimed at restricting the application of the civil offence to factual situations which demonstrate some *intent to cause serious harm*. Certainly the Discussion Paper contrasts the proposed regulation of adult cyber-abuse material with the lower thresholds set for cyber-abuse material directed at children. It states:

> The focus of the new cyber abuse scheme would be on serious cases of abuse, recognising that adults can be expected to demonstrate a higher level of resilience and maturity than children, and that it will be important to avoid creating an unreasonable regulatory burden.

[17] *Trkulja v Google LLC* (2018) 263 CLR 149, [31].
[18] Ibid [32].
[19] *Favell v Queensland Newspapers Pty Ltd* (2005) 79 ALJR 1716, [26].

> The cyber abuse scheme would aim to provide a safeguard for serious instances of online harassment and humiliation...

As regards s7(1)(b) specifically, the Discussion Paper notes:

> This definition is intended to set a higher threshold for what constitutes adult cyber abuse compared with the cyberbullying of an Australian child. For adults, the material in question would need to be intended to have an effect of causing **serious distress or harm**, rather than intended to have **an effect on the person**.

However, it is notoriously difficult to prove a person's subjective intent. Hence, the test seeks to sever the Gordian knot of these two competing considerations, retaining the threshold for establishing an intention to cause harm, but by considering whether it is *likely* that the material discloses such an intention when considered by the ordinary, reasonable person, rather than enquiring as to the mental state of the particular person who posted the material.

While this test may well avoid the difficulty of establishing a person's subjective intention, it introduces a different problem. If the test articulated in s7(1)(b) is constructed in the manner in which it operates in defamation law, it will have the effect of focusing inquiry on the *hypothetical* question of what intention likely sat behind the material, construed by reference to the material itself and not to the evidence taken as a whole. That is – in an extreme case – there might be direct evidence that the person publishing the material *in fact* intended to cause harm. However, if their methods are subtle and especially if they are personal, as they so often are in instances of cyber-abuse, we might plausibly imagine a scenario where that the intention is not clear on the face of the material. In such a case, the other direct evidence as to their intention would not be relevant, and a case against them would fail for want of establishing that the ordinary, reasonable person would consider that the material likely disclosed an intention to cause serious harm.

This construction of the test is not fanciful. Rather, it flows as a matter of proper statutory interpretation from contrast with the element at subsection 7(1)(c). That element requires establishing that:

> (c) an ordinary reasonable person in the position of the Australian adult would regard the material as being, in all the circumstances, menacing, harassing or offensive;

In that test the objective assessment is made with regard to the material 'in all the circumstances'. By implication the lack of this phrase in the s7(1)(b) test suggests that its application ought not to consider 'all the circumstances'. As well as difficult which arises from potentially excluding direct evidence of the offender's subjective intention, the difference between the tests in ss7(1)(b) and (c) gives rise to possibility of inconsistent outcomes. That is, material might only be 'menacing, harassing or offensive' when it is considered in all of the circumstances. However, inquiry as to whether that material was likely

intended to cause harm then, *prima facie*, excludes those circumstances from consideration. If it is only possible to establish that the material is 'menacing, harassing or offensive' with regard to those circumstances, it seems to follow that establishing that that same material was likely intended to cause serious harm will require considering the same matters. Excluding those circumstances from the inquiry in s7(1)(b) seems odd on its face.

There are two methods by which this difficulty might be remedied. The first route is to amend the s7(1)(b) test to make explicit that 'all the circumstances' are relevant when making a finding as to the likely intent. However, it is not clear that an intent element is necessary in the circumstances. The civil offence already requires a finding, per s7(1)(c), that the material posted was 'menacing, harassing or offensive'. Section 8 already contains what effectively amounts to a 'public interest test' when it provides:

> **8 Determining whether material is offensive**
>
> (1) The matters to be taken into account in deciding for the purposes of this Act whether an ordinary reasonable person in the position of a particular Australian adult would regard particular material as being, in all the circumstances, offensive, include:
>
> (a) the standards of morality, decency and propriety generally accepted by reasonable adults; and
>
> (b) the literary, artistic or educational merit (if any) of the material; and
>
> (c) the general character of the material (including whether it is of a medical, legal or scientific character).

Given this safeguard, it may not be necessary to interrogate whether a person held a particular intention when posting material that is menacing, harassing or offensive. Analogy might be taken to the offence contained in s78 of the *Criminal Code 1913* (WA), which establishes liability for 'any conduct, otherwise than in private, that is likely to create, promote or increase animosity towards, or harassment of, a racial group, or a person as a member of a racial group'. That offence – criminal, not civil, and carrying a potential maximum sentence of 5 years imprisonment – operates to attach criminal fault with strict liability. Like the proposed s7 offence, it already requires establishing the likely consequence of the conduct and is modified by a broad public interest test, s80G which provides relevantly:

> **80G. Defences to s. 78, 80, 80B or 80D charge**
>
> (1) It is a defence to a charge under section 78 or 80B to prove that the accused person's conduct was engaged in reasonably and in good faith
>
> (a) in the performance, exhibition or distribution of an artistic work; or

(b) in the course of any statement, publication, discussion or debate made or held, or any other conduct engaged in, for

(i) any genuine academic, artistic, religious or scientific purpose; or

(ii) any purpose that is in the public interest; or

(c) in making or publishing a fair and accurate report or analysis of any event or matter of public interest.

For the reasons articulated above, particularly the difficulties relating to tests for intention, it might be preferable to deal with the issue it raises by a second route: removing s7(1)(b) entirely, and instead expanding s8 to include a more robust public interest test, drawn from language such as that in s80G of the *Criminal Code 1913* (WA). Hence our recommendation:

4. Section 7(b) of the proposed legislation be amended either to remove the test of intention and provide for a strict liability offence, or to clarify that 'all the circumstances' related to the material are relevant to interrogation as to the likely intention.

7.2 The Definition of Intimate Image

The Discussion Paper notes that 'It is not proposed to substantively change the operation of the image-based abuse scheme. The scheme is modern, has appropriate coverage of services, and is operating effectively.' We agree with this assessment. However, question 17 asks:

17. Does the image-based abuse scheme require any other modifications or updates to remain fit for purpose?

There is room for modification of the scheme to clarify precisely when an image will fall within the definition of 'intimate'. It is not expected that the recommendation below will be a substantive change to the operation of the scheme. Rather it will avoid potential edge cases from being outside the operation of that scheme.

The draft Bill proposes to re-establish the civil offence of posting an intimate image, which is presently captured by the *Enhancing Online Safety Act 2015* (Cth). The relevant civil offence is set out in the draft bill at section 75:

75 Posting an intimate image

(1) A person (the first person) must not post, or make a threat to post, an intimate image of another person (the second person) on:

(a) a social media service; or

(b) a relevant electronic service; or

(c) a designated internet service;

if:

(d) the first person is ordinarily resident in Australia; or

(e) the second person is ordinarily resident in Australia.

Civil penalty: 500 penalty units.

Consent

(2) Subsection (1) does not apply if the second person consented to the posting of the intimate image by the first person...

The definition of intimate image is given by s15, which breaks it into three main categories: depiction of private parts; depiction of private activity; and depiction of person without attire of religious or cultural significance.

An image can be 'intimate' because it is a depiction of private parts per s15(2) if:

(a) the material consists of a still visual image or moving visual images; and

(b) the material depicts, or appears to depict:

(i) the person's genital area or anal area (whether bare or covered by underwear); or

(ii) if the person is female or a transgender or intersex person identifying as female–either or both of the person's breasts;

in circumstances in which an ordinary reasonable person would reasonably expect to be afforded privacy.

Despite some scholarly commentary on this provision as it exists in the *Enhancing Online Safety Act 2015* (Cth), there is, as yet, no case law on interpretation of the provision. Nor does the scholarly commentary address the question of construction of the definition of 'intimate image'.[20]

Two potential difficulties arise with construction of this section. The first is how the conditional clause in s15(2)(b)(ii) should be read. One reading suggests that the condition (if the person is female or a transgender or intersex person identifying as female) should be read disjunctively from the clause. Such a reading would mean that s15(2)(b)(ii) is triggered only if the person who is subject of the image meets those criteria. The alternative reading is that the condition should be read conjunctively with the clause. On this reading material

[20]See, for example Michelle Evans, 'Regulating the Non-Consensual Sharing of Intimate Images (Revenge Pornography) via a Civil Penalty Regime: A Sex Equality Analysis' (2018) 44 *Monash University Law Review* 602–620 <https://heinonline.org/HOL/P?h=hein.journals/monash44&i=602>.

would be captured by the definition if that image appears to depict a person who is female or a transgender or intersex person identifying as female. The former reading privileges the experience/evidence of the person who is the subject of the picture. The latter reading privileges the evidence of the hypothetical viewer of the image.

As a matter of statutory construction the former interpretation is likely to be preferred by a court. Specifically, it is difficult to give effect to the phrase 'identifying as' unless that section interrogates the image subject's actual intention. However, even if this is the correct interpretation, then another difficulty follows; it is not clear how the section should be interpreted if a person changes their gender identification, nor when a person does not identify as any particular gender.

To illustrate the first difficulty, consider an image is taken at a particular point in time which depicts either or both of a person's breasts. If at some later point in time that person transitions from female-identifying to no longer female-identifying, would the earlier picture lose its status as 'intimate'? That is, it is not clear whether the definition applies at the moment the image is created, or at some later point in time. One interpretation of s15(2)(b)(ii) is that someone target a person who was undergoing such a transition with the intention of taking and distributing what would otherwise be 'intimate' photos of them topless, but rely on the fact of their no longer identifying as female to avoid the application of the definition of intimate. Alternatively, a malicious actor might target someone who was non-binary identifying for distribution of their intimate images, specifically. One can certainly imagine a person who was intersex and did not identify as female who nevertheless would want to maintain that a picture depicting their breasts was relevantly 'intimate'. However, the construction of the test identified above would exclude application of the definition from persons who were non-binary identifying.

Certainly, s15(3)(b)(i), contains a broader definition of an intimate image, providing that an image will be intimate where:

> (a) the material consists of a still visual image or moving visual images; and
>
> (b) the material depicts, or appears to depict, the person:
>
> (i) in a state of undress;
>
> ...
>
> in circumstances in which an ordinary reasonable person would reasonably expect to be afforded privacy.

However, it is not clear that a broad interpretation of s15(3) would resolve the issue. Specifically, the rule against surplusage requires that an interpretation of s15(2) be given which which does not render it otiose.[21]

Appreciating that the proposed language is drawn from legislative provisions which are widely enacted, there are nevertheless reasons for amending this section. The difficulties with statutory construction noted above are real. Unless

[21] See generally, *Project Blue Sky v Australian Broadcasting Authority* (1998) 194 CLR 355.

these matters are clarified by the legislature, they will require construction by a court. In practical terms this means that a person who – but for these issues – would have been a victim of an offence under s15 will have to act as a 'test case' for the proper construction of that section. It would be preferable to avoid that situation by clarifying how the test is to be applied by taking the opportunity presented to redraft it.

Hence, our recommendation:

6. Section 15(2) be redrafted to simplify and clarify the application of that section.