Facebook submission to the review of model defamation provisions

(STAGE TWO)

31 MAY 2021

FACEBOOK

Executive Summary

Facebook welcomes the opportunity to provide comments in response to the Council of Attorneys-General's (CAG) discussion paper *Review of Model Defamation Provisions – Stage 2*. We have supported CAG's work to reform Australian defamation law, including considering how to make defamation law more fit-for-purpose in the digital age. We welcomed Stage 1 and look forward to contributing to Stage 2.

The Discussion Paper outlines its task as *"to consider what liability intermediaries should have in defamation – based on the extent to which they contribute to the risk of harm to reputation resulting from the publication of user-generated content. This requires an understanding of what internet intermediaries do and are capable of doing."* We agree.

As our CEO Mark Zuckerberg said in October 2019, "People having the power to express themselves at scale is a new kind of force in the world — a Fifth Estate alongside the other power structures of society. People no longer have to rely on traditional gatekeepers in politics or media to make their voices heard, and that has important consequences."[1]

Millions of people in Australia choose to take advantage of technology and use their voice for self-expression, to advocate for social causes, or encourage political change. However, we know others may use their voice in ways that are problematic, including to defame others.

The role of internet intermediaries in relation to defamatory content deserves further examination by Australian policymakers. The discussion paper contains thoughtful and welcome discussion about the need to clarify and strengthen intermediary liability around defamatory content. We strongly support the discussion paper's position that internet intermediaries should not automatically be held liable for the contents of material that is authored or created by a third party and shared on their platform. Recent cases have created significant uncertainty for internet intermediaries by essentially requiring them to block access to content, solely on the basis of a user allegation that the content is defamatory.[2]

Greater clarity about intermediaries' ability to rely on existing defences would reduce the risk of over-blocking and better protect free expression. It also ensures the author or creator of the material - the party with the greatest level of control over the

---

[1] Facebook, 'Mark Zuckerberg Stands For Voice and Free Expression', *Facebook Newsroom*, 17 October 2019, https://about.fb.com/news/2019/10/mark-zuckerberg-stands-for-voice-and-free-expression/.
[2] Defteros v Google LLC [2020] VSC219

material's contents - rightly continues to bear primary responsibility and therefore liability for defamation.

However, we also recognise the concerns raised in the discussion paper that defamation law should enable efficient resolution for individuals who are the subject of defamatory material.

If internet intermediaries' protection against liability for unlawful defamatory content is strengthened and clarified, we recognise that policymakers and the community may have greater confidence in the responsiveness of digital platforms by simultaneously making that liability protection conditional on companies' ability to meet best practices to combat the spread of that content. Instead of granting internet intermediaries an automatic, blanket immunity, an immunity for secondary publishers could be contingent on the intermediary having adequate systems in place, such as participation in an appropriate complaints process (connecting complainants with primary publishers) or systems to restrict a piece of content in Australia, after a sufficiently rigorous independent process to determine that the content is defamatory under Australian law.

In assessing the various options for defences and immunities raised in the discussion paper, we have been guided by the following principles:

- There should be rigorous independent processes for assessing whether material is defamatory. Digital platforms should not be required to make their own assessments about whether material is defamatory: we are not in a position to accurately make those assessments, given we could only rely on unverified and potentially conflicting claims by opposing parties.

- Even though courts have an essential role to play in deciding on matters of defamation, we acknowledge the discussion paper's perspective that it may not be efficient to rely entirely on court orders before restricting defamatory material.

- The legislation should not inadvertently encourage digital platforms to excessively censor or restrict material. It is possible for online material to generate significant public benefit, while also being potentially defamatory. Recent years have seen many significant social movements, including: #MeToo and discussions about women's safety and economic security; the injustices highlighted by the Black Lives Matter movement or the campaign around Aboriginal Deaths in Custody; or whistleblowers for systemic child sexual abuse occurring within institutions. All these movements have been based on individuals shedding light on critical social issues, and in some cases making

allegations of criminal conduct against individuals in positions of power - in particular, via social media services like Facebook. For this reason, we allow people to use our services to level allegations of criminal conduct against other individuals.

In our experience, many of the individuals subject to such complaints have claimed (rightly or wrongly) that this material is defamatory. It is incredibly challenging to adjudicate these claims. If defamation legislation does not provide sufficient clarity for digital platforms on how to consider claims in instances such as this, it would instead encourage digital platforms to excessively censor or restrict material in order to manage potential liability under the legislation.

Of the four options for defamation law reform proposed in the paper, Facebook's preference is for a hybrid model that draws from two of the models proposed in the paper. We have consistently supported amendments to the innocent dissemination defence to clarify its application to internet intermediaries[3], including by introducing a rebuttable presumption that internet intermediaries are subordinate distributors. We believe this approach is proportionate and would provide greater clarity about the requirements expected of internet intermediaries.

However, to address concerns raised in the discussion paper about efficient resolution for individuals who are the subject of defamatory material, we suggest this innocent dissemination option could be enhanced if it was accompanied by two other changes to benefit individuals who are raising complaints about defamatory material:

1. A series of amendments to the pre-trial processes, in particular, amending the concerns notice scheme to better reflect the limited role that internet intermediaries can play in this process; and

2. The introduction of a safe harbour inspired by section 5 of the Defamation Act in the United Kingdom but substantially streamlined to address concerns about the practicality of this regime. This safe harbour could ultimately require internet intermediaries to assist individuals to identify the author or creator of potentially defamatory material, so that the complainant can pursue redress against the author or creator directly. While we support the principle of section 5, the current procedural steps required to be taken by a website operator are operationally burdensome, and place the website operator in the awkward position of acting as a go-between for the complainant and the operator. A more effective regime could require internet intermediaries to assist the

[3] Facebook, *Facebook's response to the review of Model Defamation Provisions*, 31 January 2020, https://www.justice.nsw.gov.au/justicepolicy/Documents/review-model-defamation-provisions-amendments/facebook-submission-to-mdaps.pdf.

complainant (in order to take advantage of the safe harbour) but then leave the two parties to proceed themselves, avoiding the assumption that content should simply be removed following an allegation that content is defamatory.

The discussion paper also seeks views on whether Australia should adopt a provision similar to section 230 in the United States' Communications Decency Act. Section 230 has created the conditions for the internet to thrive, and for internet intermediaries to enable billions of people to express themselves online; however, Facebook has also called for thoughtful amendments to make it work better. This submission is therefore not arguing for wholesale adoption of a section 230-type provision in Australian defamation law but rather acknowledges that the extension of an appropriate immunity or safe harbour to internet intermediaries has strong benefits for the management of online content.

While we support the need for reforming defamation law, it is worth noting that the discussion paper's proposed options are based on a misunderstanding about Facebook's approach to addressing harmful content. The discussion paper mischaracterises Facebook's business incentives by suggesting that algorithms are used by digital platforms to profit from harmful content, and that legislative reform is required to enact a "reconsideration of business models to better protect users from the risk of harm to reputation."

We have strong business incentives to limit the spread of sensational or harmful content on our platform, and we invest heavily in systems to assist in detecting this content via artificial intelligence or user reports.[4] Facebook's algorithms do not promote or amplify harmful content: as outlined in a recent essay by Facebook's vice-president for global affairs Nick Clegg, ranking algorithms are simply a method for prioritising content that users have already opted to see.[5] The use of algorithms does not have any bearing on the level of editorial control over the contents of published material, and so it should not have any relation to liability under defamation law.

While Facebook does not have editorial control over the content distributed on our services, we invest very significantly in policies, processes and products to combat the spread of harmful content. Facebook is subject to Australian defamation laws and, accordingly, if a user reports content for defamation and an Australian court has found the content at issue to be unlawfully defamatory, Facebook will disable access to that content in Australia. We have also released products to assist users to limit the

---

[4] Further information is available in our Community Standards Enforcement Report, available at: G Rosen, 'Community Standards Enforcement Report, First Quarter 2021', *Facebook Newsroom*, https://about.fb.com/news/2021/05/community-standards-enforcement-report-q1-2021/, 19 May 2021.

[5] N Clegg, 'You and the Algorithm: It Takes Two to Tango', *Medium*, 31 March 2021, https://nickclegg.medium.com/you-and-the-algorithm-it-takes-two-to-tango-7722b19aa1c2.

risk of defamatory material on their Page or Group: for example, in March 2021, we released a new product called Control Who Can Comment, which allows the admin of a Facebook Page to limit who can comment on particular posts. Admins are able to designate posts that no user can comment on.[6] The ability to limit comments was made in response to feedback from a range of users and stakeholders, including Australian news publishers, and this change directly empowers Facebook Page admins to manage third party publications and potential liability arising from decisions like *Voller*.

Facebook stands ready to continue contributing to the debate about models of defamation law reform that could adapt the law to be more suitable for the digital age. We welcome the opportunity to continue to invest in policies, procedures and products that ensure the restriction of access to harmful defamatory material in Australia, out of respect for Australian law, while also allowing people the many benefits of having the power to express themselves at scale.

[6] R Sethuraman, 'More Control and Context in News Feed;, *Facebook Newsroom*, 31 March 2021, https://about.fb.com/news/2021/03/more-control-and-context-in-news-feed/.

Table of contents

EXECUTIVE SUMMARY **2**

INTRODUCTION **8**

SPECIFIC COMMENTS ON ISSUES IN THE DISCUSSION PAPER **12**

Issue 1: Categorising internet intermediaries **12**

Issue 2: Immunities and defences **13**
- Option 1 17
- Option 2 18
- Option 3 19
- Option 4 21

Issue 3: Complaints notice process **22**

Issue 4: Power of courts to order that material be removed **24**

Issue 5: Power of courts to order that internet intermediaries reveal the identity of originators **25**

Other issues **26**

Introduction

Facebook welcomes the opportunity to provide comments to the Review of Model Defamation Provisions. We commend the Council of Attorneys-General (CAG) on the review to adapt Australia's defamation laws for the digital age. We support the changes in Stage 1 of the reforms, and welcome the opportunity to contribute to the public debate and discussion around potential measures in Stage 2 of the reforms.

As our CEO Mark Zuckerberg said in October 2019, "People having the power to express themselves at scale is a new kind of force in the world — a Fifth Estate alongside the other power structures of society. People no longer have to rely on traditional gatekeepers in politics or media to make their voices heard, and that has important consequences."[7]

Millions of people in Australia choose to take advantage of technology and use their voice for self expression, to advocate for social causes or encourage political change. However, we know others may use their voice in ways that are problematic, including to defame others.

The role of internet intermediaries in relation to defamatory content deserves further examination by Australian policymakers. In our last submission on defamation law reform[8], we expressed our view that Australian laws are in many ways not fit-for-purpose in the digital world.

The discussion paper contains thoughtful and welcome discussion about the need to clarify and strengthen intermediary liability around defamatory content. We agree with the position that we should not be automatically held liable for the contents of material that is authored or created by a third party and shared on our platform. If platforms were made liable for that material instead of the author or creator, it would potentially perversely create an incentive for the author or creator to publish more defamatory material, because they would no longer be the party who is primarily liable.

If internet intermediaries' protection against liability for unlawful defamatory content is strengthened and clarified, we recognise that policymakers and the community may have greater confidence in the responsiveness of digital platforms by simultaneously making that liability protection conditional on companies' ability to meet best

---

[7] hFacebook, 'Mark Zuckerberg Stands For Voice and Free Expression', *Facebook Newsroom*, 17 October 2019, https://about.fb.com/news/2019/10/mark-zuckerberg-stands-for-voice-and-free-expression/.

[8] Facebook, *Facebook's response to the review of Model Defamation Provisions*, 31 January 2020, https://www.justice.nsw.gov.au/justicepolicy/Documents/review-model-defamation-provisions-amendments/facebook-submission-to-mdaps.pdf.

practices to combat the spread of that content. Instead of granting internet intermediaries an automatic, blanket immunity, an immunity for secondary publishers could be contingent on the intermediary having adequate systems in place, such as participation in an appropriate complaints process (connecting complainants with primary publishers) or systems to restrict a piece of content in Australia, after a sufficiently rigorous independent process to determine that the content is defamatory under Australian law.

While Facebook does not have editorial control over the content distributed on our services, we invest very significantly in policies and processes to combat the spread of harmful content. Facebook is subject to Australian defamation laws and, accordingly, if a user reports content for defamation and an Australian court has found the content at issue to be unlawfully defamatory, Facebook will disable access to that content in Australia.

We have released products to assist the administrators of Facebook Pages or Facebook Groups to limit the risk of defamatory material on their Page or Group. For example, in March 2021, we released a new product called Control Who Can Comment, which allows the admin of a Facebook Page to limit who can comment on particular posts. Admins are able to designate posts that no user can comment on.[9] This product change was made in response to feedback from a range of users and stakeholders, including Australian news publishers. As with many others, we await the High Court's decision on whether it will uphold the "curious" decision[10] taken in the *Voller* case.

The significant amount of work that we have done to combat the spread of harmful content demonstrates that the discussion paper does not correctly describe our incentives: we have a strong business incentive to limit the spread of sensational or harmful content on our platform, and we invest heavily in systems to assist in detecting this content via artificial intelligence or user reports.

The Discussion Paper does not correctly describe the way in which Facebook and Instagram's services work when it states: "Social media platforms generally display content for consumption as linear 'feeds', curated by algorithms or displayed chronologically. Examples include Facebook, Instagram and Snapchat. Platforms may also offer additional functions, including instant messaging services."[11] On Facebook

---

[9] R Sethuraman, 'More Control and Context in News Feed;, *Facebook Newsroom*, 31 March 2021, https://about.fb.com/news/2021/03/more-control-and-context-in-news-feed/.

[10] C Porter Address to the National Press Club, https://www.attorneygeneral.gov.au/media/speeches/address-national-press-club-canberra-20-november-2019, November 2019.

[11] Page 37. The Discussion Paper relies on the Australian Competition and Consumer Commission's Digital Platforms Inquiry, but it refers to a position that was not covered in the DPI Report. In addition,

and Instagram, this is not accurate. People choose who to friend, what Pages or accounts to like, and what Groups to join. Ranking and prioritisation algorithms provide only the content that people have opted into, in the most relevant order for that individual. These algorithms make it easier for people by ensuring the time that they spend on our services is useful and valuable to them.

Similarly, the discussion paper misunderstands our business model by claiming it is based on the promotion or amplification of harmful content. As outlined in a recent essay by Facebook's vice-president for global affairs Nick Clegg, ranking algorithms are simply a method for prioritising content that users have opted to see.[12] The use of algorithms does not have any bearing on the level of editorial control over the contents of published material, and so it should not have any relation to liability under defamation law.

We believe that, as policymakers are considering what processes are sufficient for identifying, assessing and restricting defamatory material, they should be guided by the following requirements:

- There should be rigorous independent processes for assessing whether material is defamatory. Digital platforms should not be required to make their own assessments about whether material is defamatory: we are not in a position to accurately make those assessments, given we could only rely on unverified and potentially conflicting claims by opposing parties.

  A number of recent cases have raised considerable uncertainty for internet intermediaries when seeking to avail themselves of relevant defences like innocent dissemination. The existing defence essentially requires internet intermediaries to remove or geo-block content (within 7 days) based only on a user allegation that the content is defamatory. This is a near impossible task to assess, but it is necessary in order to preserve relevant defences. In practice, this means a more chilling effect on free expression, where allegations of defamatory content online are considered true, without being tested or assessed by a court.[13]

- Even though courts have an essential role to play in deciding on matters of defamation, we acknowledge the discussion paper's perspective that it may not be efficient to rely entirely on court orders before restricting defamatory material. We recognise the concerns raised in the discussion paper about the

---

many sections of the DPI Report are based on misconceptions or no evidence, and we recommend that the Attorneys-General do not rely on the findings of the DPI Report for this reform.

[12] N Clegg, 'You and the Algorithm: It Takes Two to Tango', *Medium*, 31 March 2021, https://nickclegg.medium.com/you-and-the-algorithm-it-takes-two-to-tango-7722b19aa1c2.

[13] For example, see *Defteros v Google LLC* [2020] VSC 219

efficiency of a system that relies entirely on individuals taking all complaints about defamatory material to courts will not enable efficient resolution for individuals. For that reason, we have provided more information below about a proposed complaints notice process that could provide, in the first instance, a more-efficient pathway for individuals to restrict access to non-contested defamatory material.

- The legislation should not inadvertently encourage digital platforms to excessively censor or restrict material. It is possible for online material to generate significant public benefit, while also being potentially defamatory. Recent years have seen many significant social movements, including: #MeToo and discussions about women's safety and economic security; the injustices highlighted by the Black Lives Matter movement and the campaign around Aboriginal Deaths in Custody; and whistleblowers for systemic child sexual abuse occurring within institutions. All these movements have been based on individuals making allegations of criminal conduct against individuals in positions of power - in particular, via social media services like Facebook. For this reason, we allow people to use our services to level allegations of criminal conduct against other individuals.

  In our experience, many of the individuals subject to such complaints have claimed (rightly or wrongly) that this material is defamatory. It is incredibly challenging to adjudicate these claims. If defamation legislation does not provide sufficient clarity for digital platforms on how to consider claims in instances such as this, it would inadvertently encourage digital platforms to excessively censor or restrict material in order to manage potential liability under the legislation.

We commend the NSW Department of Communities and Justice for the thoughtful and welcome proposals outlined in the discussion paper. There could be a number of potential policy reform options that may be able to meet these objectives. And so, to assist policymakers in considering ways forward, Facebook has identified a preferred option that we believe would be most effective - but the remainder of this submission contains discussion about both the advantages and the disadvantages of each option in the discussion paper.

Specific comments on issues in the discussion paper

Issue 1: Categorising internet intermediaries

The Discussion Paper states that "*the task of this Discussion Paper is to consider what liability intermediaries should have in defamation – based on the extent to which they contribute to the risk of harm to reputation resulting from the publication of user-generated content. This requires an understanding of what internet intermediaries do and are capable of doing.*"[14]

Based on an academic paper that was subsequently published as a book in the UK, the CAG asks whether it is appropriate to group the functions of internet intermediaries,[15] and their consequent liability, into three categories: basic Internet services, digital platforms and forum administrators. [16]

The proposed functions-based approach to categorising internet intermediaries has some inherent flaws, including:

- categorising internet intermediaries based on their functions is not technology neutral, and risks being quickly outdated as technology develops and the functions of internet intermediaries continue to evolve;
- many internet intermediaries have multiple and distinct functions and features, and the Discussion Paper does not identify whether these functions would be assessed and categorised individually, or as an amalgamated whole. Neither approach is practical as:
    - contrary to the Discussion Paper's suggestion that a functions-based approach would offer certainty,[17] categorising internet intermediaries by individual functions risks creating too much uncertainty about how internet intermediaries would be categorised in any given case; and
    - categorising internet intermediaries based on an amalgamation of their functions risks losing important distinctions between different functions which, if taken individually, may have been classified differently. For example, a search engine may offer a marketplace-style system for reviewing businesses and services embedded within their search product.

---

[14] Discussion Paper at [3.24]

[15] At [3.25] the Discussion Paper adopts the definition of 'internet intermediaries' by the OECD, namely entities that "*bring together or facilitate transactions between third parties on the Internet. They give access to, host, transmit and index content, products and services originated by third parties on the Internet or provide Internet-based services to third parties.*"

[16] Discussion Paper at [3.27].

[17] Discussion Paper at [3.88].

Further, categorising internet intermediaries for the purpose of determining liability would place Australia further out of step with other jurisdictions, for which applicable defences/immunities apply broadly to providers of interactive computer services,[18] website operators,[19] and information society service providers.[20]

As outlined in the Discussion Paper, the question of whether a defendant is a publisher is a matter for the evidence in each case. Creating categories of internet intermediaries based on certain functions is neither helpful nor necessary for defamation reforms addressing online publication. Existing legal principles which apply to secondary publishers, centring on elements like participation and control[21] -- together with necessary changes to the defence of innocent dissemination (which we outline later in this submission) -- will properly address the question of the responsibility of internet intermediaries in the publication of defamatory content. These same principles should apply to all secondary publishers - including digital platforms and forum administrators.

As a result, we believe further categorisation of internet intermediaries (beyond the existing characterisation between primary and secondary publishers) is not necessary.

Issue 2: Immunities and defences

Internet intermediaries are not the authors or creators of content and are not in a position to assess whether user-generated content is unlawfully defamatory. If internet intermediaries are expected to make these assessments, they will have few other options than to be overly-broad in the content that they block or restrict to avoid liability -- an outcome that would be detrimental for free expression.
Not every complainant's assertion that content is defamatory is correct. For example, when content raises allegations of criminal conduct, the complainant has a vested interest in removing this material, especially if it is true.

Assessments about whether content is defamatory are more appropriately resolved with the originator, or via rigorous independent third parties (like the courts), in order to consider the precise reasons for publication and whether valid defences arise. Facebook does make significant investments and take action against whole categories of harmful content, which may also incidentally be defamatory. Content on Facebook is subject to our Community Standards - a comprehensive set of principles that govern what content is and is not permitted on Facebook.[22] For example, our

---

[18] Section 230, *Communications Decency Act 1996* (US).
[19] Section 5, *Defamation Act 2013* (UK).
[20] Articles 14 and 15 of the *EU Directive on Electronic Commerce (2000/31/EC).*
[21] [Thompson v Australian Capital Television Pty Ltd [1996] HCA 38]
[22] Facebook, *Community Standards*, www.facebook.com/communitystandards

Community Standards prohibit content that is bullying, harassment and hate speech (which includes content that directly attacks people based on their race, ethnicity, national origin, religious affiliation, sexual orientation, sex, gender or gender identity and serious disabilities or diseases). When applying these policies, we seek to find a balance between safety and allowing users to have a voice. To show the scale of content that is removed and the emphasis we place on safety and reducing harmful content on our services, in the period January to March 2021:

- 8.8 million pieces of content were removed for bullying and harassment, 54.4 per cent of which was detected proactively via artificial intelligence and
- 25.2 million pieces of content were removed for hate speech, 96.8 per cent of which was detected proactively via artificial intelligence.[23]

Assessing whether content is defamatory under Australian law is more challenging. We do not have the context that the originator of the content has to properly assess the strength of the claim or any other defence. Accordingly, any content that a complainant alleges is defamatory, and where it otherwise does not violate our Community Standards, is highly likely to be nuanced and relevant to precise facts and circumstances that are only within the knowledge of the originator of the content-at-issue and the complainant. A rigorous independent process is the best method for determining whether the content is defamatory, rather than asking private companies to undertake assessments of the facts and make their own determination.

The Discussion Paper sets out 4 options for reform to clarify or modify the liability of internet intermediaries for third party content, by:

- Option 1: retaining the status quo with some minor changes to the Model Defamation Provisions (MDPs) to clarify the role of internet intermediaries, by:

  Option 1a: [24] maintaining the existing approach, whereby subject to any defences and the facts of an individual case, an internet intermediary can be found to be the publisher of third-party content published using their services, and pursuant to which it is currently unclear whether a social media platform is considered a publisher under Australian defamation law;

  Option 1b: [25] introducing statutory immunity for 'basic internet services' from defamation liability, on the basis that they are 'mere conduits' that do not

---

[23] G Rosen, 'Community Standards Enforcement Report - February 2021', Facebook Newsroom, 11 February 2021, https://about.fb.com/news/2021/02/community-standards-enforcement-report-q4-2020/

[24] Discussion Paper at [3.70 – 3.81].

[25] Discussion Paper at [3.82 – 3.91].

actively participate in the publication or sufficiently contribute to the risk of reputational harm;

Option 1c: [26] amending the current concerns notice and/or offer to make amends process in the MDPs to better apply to internet intermediaries (for example, by removing the mandatory requirements that any offer to make amends include an offer to publish a correction or clarification) and thus better enable them to rely upon the defence of a reasonable offer to make amends.

- Option 2: clarifying the innocent dissemination defence in clause 32 of the MDPs in relation to digital platforms and forum administrators, by:

  Alternative A: [27] creating a default position that digital platforms and forum administrators are not primary distributors; or

  Alternative B: [28] introducing a presumption that digital platforms and forum administrators are subordinate distributors. This presumption could be rebuttable (for example, if the complainant shows the platform/administrators adopted, curated, promoted the content at issue), and could specify what constitutes "notice" to clarify when the innocent dissemination defence applies.

- Option 3:[29] introducing a safe harbour similar to section 5 of the *Defamation Act 2013* (UK) (UK Act) for internet intermediaries for user-generated content, subject to a complaints notice process; or

- Option 4:[30] introducing immunity similar to section 230 of the *Communications Decency Act 1996* (US) (CDA) [31] for all internet intermediaries for user-generated content unless the internet intermediary materially contributes to the unlawfulness of the publication.

Of the four options for defamation law reform proposed in the paper, Facebook's preference is for a hybrid model that draws from two of the models proposed in the paper. We have consistently supported amendments to the innocent dissemination

[26] Discussion Paper at [3.92 – 3.106].
[27] Discussion Paper at [3.116 – 3.117].
[28] Discussion Paper at [3.118 – 3.126].
[29] Discussion Paper at [3.127 – 3.146].
[30] Discussion Paper at [3.147 – 3.164].
[31] Section 230(c)(1) CDA provides: "*No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider*".

defence (Option 2b) to clarify its application to internet intermediaries[32], including by introducing a rebuttable presumption that internet intermediaries are subordinate distributors. We believe this approach is proportionate and would provide greater clarity about the requirements expected of internet intermediaries.

However, to address concerns raised in the discussion paper about efficient resolution for individuals who are the subject of defamatory material, we suggest this innocent dissemination option could be enhanced if it was accompanied by two other changes to benefit individuals who are raising complaints about defamatory material:

1. A series of amendments to the pre-trial processes, in particular, amending the concerns notice scheme to better reflect the limited role that internet intermediaries can play in this process (Option 1c); and

2. The introduction of a safe harbour inspired by section 5 of the Defamation Act in the United Kingdom (Option 3) but only if it were substantially streamlined to address concerns about the practicality of this regime. This safe harbour could ultimately require internet intermediaries to assist individuals to identify the author or creator of potentially defamatory material, so that the complainant can pursue redress against the author or creator directly. While we support the principle of section 5, the current procedural steps required to be taken by a website operator are operationally burdensome, and place the website operator in the awkward position of acting as a go-between for the complainant and the operator. A more effective regime could require internet intermediaries to assist the complainant (in order to take advantage of the safe harbour) but then leave the two parties to proceed themselves, avoiding the assumption that content should simply be removed following an allegation that content is defamatory.

In relation to Option 4, section 230 has created the conditions for the internet to thrive, and for internet intermediaries to enable billions of people to express themselves online; however, Facebook has also called for thoughtful amendments to make it work better. This submission is not arguing for wholesale adoption of a section 230-type provision in Australian defamation law but rather an acknowledgement that the extension of an appropriate immunity or safe harbour to internet intermediaries has strong benefits for the management of online content.

---

[32] Facebook, *Facebook's response to the review of Model Defamation Provisions*, 31 January 2020, https://www.justice.nsw.gov.au/justicepolicy/Documents/review-model-defamation-provisions-amendments/facebook-submission-to-mdaps.pdf.

Option 1

We agree with the measures introduced following Stage 1 of this review process, which update pre-trial procedures to better reflect the position of internet intermediaries and online publishing, including by introducing: a mandatory concerns notice regime, requirements that a valid concerns notice specify the location of the allegedly defamatory matter (e.g. by URL for websites) and describe the serious harm suffered by the complainant. While these measures go some way to adapting these pre-trial procedures for online publication, we support the proposal in Option 1c to further update these processes.

Option 1c is consistent with revisions that we proposed in the previous Facebook submission, in which we supported the now approved draft amendments to the pre-trial procedures described above, and advocated for further amendments to these procedures to distinguish between the different positions of primary publishers and secondary publishers.[33]

To illustrate these different positions, we submitted that a valid offer to make amends requires that the publisher offer to publish, or join in publishing, a correction and pay the reasonable expenses of the aggrieved person. While a primary publisher (for example, a social media user who is an author or creator) is in the best position to determine whether content is untrue or otherwise warrants a correction, the secondary publisher (for example, an internet intermediary) is generally not in a position to determine whether the content is true or whether other defences may apply, and is not in a position to make an informed decision about whether to make an offer. Accordingly, the pre-trial process, and the associated defences,[34] are of little utility to secondary publishers like Facebook.

The Discussion Paper's consideration of Option 1(c),[35] notes that amending this process (or creating a separate process for internet intermediaries) could "*address the potential shortfalls of the current process when it comes to complaints of a third-party*".

We support amendments in line with Option 1c, to remove existing barriers to internet intermediaries participating in pre-litigation procedures, specifically, by removing (or

---

[33] Facebook, *Facebook's response to the review of Model Defamation Provisions*, 31 January 2020, https://www.justice.nsw.gov.au/justicepolicy/Documents/review-model-defamation-provisions-amendments/facebook-submission-to-mdaps.pdf.

[34] Where an offer to make amends (i) is accepted and carried out, the aggrieved person cannot assert, continue, or enforce an action for defamation in relation to the matter in question, or (ii) is not accepted, it is a defence to an action for defamation against the publisher if the offer was reasonable, made within the required time, and the publisher was ready and willing to carry out its terms (clauses 17 and 18 MDPs)

[35] Discussion Paper at [3.100].

making voluntary) the following mandatory requirements for a valid offer to make amends:[36]

- inclusion of an offer to publish, or join in publishing, a reasonable correction of the matter in question or, if the offer is limited to any particular defamatory imputations, the imputations to which the offer is limited, and
- if material containing the matter has been given to someone else by the publisher or with the publisher's knowledge—inclusion of an offer to take, or join in taking, reasonable steps to tell the other person that the matter is or may be defamatory of the aggrieved person, and
- inclusion of an offer to pay the expenses reasonably incurred by the aggrieved person before the offer was made and the expenses reasonably incurred by the aggrieved person in considering the offer.

Such revisions should be made irrespective of whether any other amendments are made to the MDPs and would remove barriers to internet intermediaries utilising the defences associated with the existing pre-trial processes.

Option 2

We support amendments to the innocent dissemination defence to better adapt it to internet intermediaries. This can be achieved by amending the current defence in clause 32 of the MDPs, or by introducing a standalone defence for internet intermediaries.

We support the proposed introduction of a presumption that an internet intermediary (which is held to be publisher) is a subordinate distributor, without otherwise having to satisfy the definition in clause 32 of the MDPs.[37] This amendment creates greater certainty for internet intermediaries, by removing the requirement that they demonstrate a lack of editorial control over the content, and instead requires the complainant to rebut the presumption.

While this amendment provides substantial clarity for internet intermediaries, the question of subordinate distributors is not the only aspect of the innocent dissemination defence that presents practical difficulties for internet intermediaries. Facebook previous submission advocated for further amendments to the innocent dissemination defence to broaden the definition of "subordinate distributor" to capture "*any person other than the author, editor, or employer of the author or editor,*

[36] Clause 15(1)(d)-(e) MDPs.

[37] Pursuant to which a person is a subordinate distributor if the person was not the first or primary distributor of the matter, the author or originator of the matter and did not have capacity to exercise editorial control over the content of the matter (or over the publication of the matter) before it was first published.

of a publication",[38] and to clarify that it is available until the subordinate distributor knows/ought to have known that the matter was unlawful (as opposed to simply knowing that the user alleges the content to be prima facie defamatory).

The existing defence essentially requires internet intermediaries wishing to preserve their innocent dissemination defence to remove or geo-block content based on nothing more than a user allegation that it is defamatory. The alternative is to not remove the content, lose the benefit of the innocent dissemination defence, and assess the risk that a merits defence (such as truth) will succeed. This decision is frequently required to be made based upon insufficient information, evidence or context to properly assess the strength of any defence. For example, an intermediary is not in a position to assess the truth or otherwise of #MeToo content, which is most appropriately resolved by a court.

We also propose further amendments to the innocent dissemination defence for internet intermediaries, specifically that the defence would require:

- that the defendant published the matter in the capacity, or as an employee or agent, of "subordinate distributor", in circumstances where internet intermediaries are presumed to be subordinate distributors unless the plaintiff can prove that they were the author, editor, or employer of the author or editor of the publication;
- the defendant did not know that the matter was defamatory, where knowledge is established upon receipt of a judgment showing that the content-at-issue is defamatory; and
- upon gaining knowledge that the content-at-issue was defamatory, the defendant failed to remove the content-at-issue within a reasonable time, where "reasonable time" aligns with the statutory timeframe for making an offer to make amends, usually 28 days from receiving sufficient information to identify the content-at-issue (ie, URL or a unique identifier).

Option 3

In principle, we support the introduction of an immunity based on section 5 of the UK Act, as:

- website operators cannot be held liable for defamatory content if the complainant could have identified (through some form of reasonable steps) the originator and brought an action directly against that person;

[38] Facebook, *Facebook's response to the review of Model Defamation Provisions*, 31 January 2020, https://www.justice.nsw.gov.au/justicepolicy/Documents/review-model-defamation-provisions-amendments/facebook-submission-to-mdaps.pdf.

- it places some onus on the complainant to try and resolve the matter directly with the originator, and if they cannot be identified, the complainant can then notify the website operator; and
- once notified, the website operator has the ability to take steps to assist the complainant and obtain the benefit of the immunity.

However, the procedural steps required to be taken by a website operator in receipt of a complaints notice under the UK approach are operationally burdensome, and place the website operator in the awkward position of acting as a go-between for the complainant and the operator.

We would only support the introduction of a defence similar to section 5 of the UK Act if it substantially streamlines the procedural aspects of the defence, including by:

- requiring the complainant to take "reasonable steps" to identify the originator, based on a clear definition of "reasonable steps" (e.g. including online and offline searches, attempting to contact the originator directly through online methods such as chat, direct message, email, etc.);

- if, by taking reasonable steps, the complainant <u>can</u> identify the originator of the content-at-issue, the internet intermediary retains immunity until the complainant obtains judgment against the originator. If the complainant <u>cannot</u> identify the originator of the content-at-issue, or the complainant has already obtained judgment against the originator, the complainant could be permitted to send a complaints notice to the relevant internet intermediary, which:
    - should have legal status (e.g. a statutory declaration) and require the complainant to provide (and confirm the correctness, on penalty for providing false information):
        - the specific content at issue by URL or other unique identifier;
        - which statements in the content they consider convey defamatory and untrue imputations (and identifying each such imputation), and why;
        - that they have a good faith belief that those statements are defamatory and not true;
        - the legal basis for the claim in Australia, including relevant statute, precedent and case law;
        - that they require the content to be removed (so that it is clear that the intermediary should handle the matter as a removal request);
        - any relevant broader context around the content being complained about;
        - how the allegedly defamatory content is causing, or is likely to cause, serious harm to the complainant; and

- the ‘reasonable steps’ the complainant has taken to identify the originator.

- An internet intermediary in receipt of a valid (i.e. compliant with the above requirements) complaints notice should be given a reasonable time to forward the complainant’s notice to the originator (if they are able to contact them) and ask for the originator’s address for service and permission to share it with the complainant. If the originator provides their address and permission, the intermediary could be required to share the originator's address with the complainant (retaining immunity until the complainant obtains a judgment against the originator); and if not, the intermediary could be required to inform the complainant of this to enable it to apply for pre-action discovery, and retains immunity until the complainant obtains a judgment against the originator. In instances where the intermediary doesn’t have information that would enable them to contact the originator, the intermediary could be required to geo-block or remove the content within a reasonable time, if they would like to retain the immunity.

Option 4

Any discussion about section 230 of the CDA should recognise upfront the tremendous benefits that this law has enabled. Section 230 has created the conditions for the Internet to thrive, and for internet intermediaries to enable billions of people to express themselves online. It provides many benefits (both in terms of clarity for internet intermediaries, and the expansion of public interest content on the internet more generally):

- The immunity (which can be lost if the provider or user of an interactive computer service “materially contributed to [the] alleged unlawfulness”) has encouraged the moderation of content by internet intermediaries. The development and enforcement of guidelines (such as Facebook’s Community Standards) has encouraged the removal of harmful content without intermediaries facing possible claims for content that they did not create;

- It relieves internet intermediaries from playing the role of judge in the assessment of whether content is unlawfully defamatory whilst retaining the ability for a complainant to seek recourse directly with the author of at-issue content.

Section 230 has been subject to much debate in the US; indeed, Facebook has argued for thoughtful changes to ensure the law is operating as it should. Given the concerns that have been raised and the consideration of reform currently underway in the US, it is not appropriate for Australia to wholesale adopt the current version of section 230.

We would welcome any discussion or consideration by Australian policymakers about whether a similar style defence could be adapted and introduced in Australia.

Issue 3: Complaints notice process

In Option 3 of the options for immunities and defence, the discussion paper proposes a safe harbour defence for an internet intermediary, if it follows a complaints notice process. This could include connecting the complainant with the originator, or remove the content. The complaints notice process is predominantly based on the current UK law.

The objective of the complaints process is to provide complainants with a way to identify the originator of defamatory content, or, where they are not identifiable (or refuse to provide their information), have that content removed or geo-blocked by the intermediary.

The Australian process for concerns notices or offering to make amends has a slightly different (although complementary) purpose: to provide a pre-trial mechanism for resolving complaints between the complainant and the publisher.

There are two points of tension between the complaints notice under the UK model, and concerns notice processes under the Australian model:

- the UK complaints notice process draws a distinction between the roles of the intermediary and the originator of allegedly defamatory content, but the AUS concerns notice process does not; and

- the AUS concerns notice process assumes that the appropriate recipient of the notice is identifiable, whereas the UK complaints notice process assumes the opposite.

There does not appear to be any advantage in introducing a complaints notice process in addition to a concerns notice process. Two schemes operating side-by-side would further complicate the process for a complainant to obtain relief, and place a double burden on both internet intermediaries and complainants.

We suggest the tensions can be addressed by absorbing aspects of the complaints notice process into a pre-litigation process for internet intermediaries. At high level, an amended pre-trial process for internet intermediaries could align with the complaints notice process described above, by requiring that:

- a complainant must take reasonable steps to identify the originator of allegedly defamatory content, in circumstances where "reasonable steps" are clearly defined;

- if, having taken reasonable steps, the complainant <u>can</u> identify the originator of allegedly defamatory content, the complainant's mandatory concerns notice must be sent to the originator, and not to the internet intermediary, in which case existing pre-litigation processes (as amended by stage 1 reforms) apply between the complainant and the originator;

- only if, having taken reasonable steps, the complainant <u>cannot</u> identify the originator of the content-at-issue, or the complainant has already obtained judgment against the originator, can the complainant's mandatory concerns notice be sent to the relevant internet intermediary;

- a mandatory concerns notice sent to an internet intermediary *must*:
  - detail the steps taken to identify the originator or provide a copy of a judgment obtained against the originator;
  - specify the location where the matter in question can be accessed (e.g. URL);
  - particularise the defamatory imputations to be relied upon, identify why they are defamatory and untrue and provide any broader context around the matter in question;
  - detail the serious harm caused to the complainant's reputation;
  - specify the relief sought (e.g. identification of the originator, removal of content, etc.);
  - provide a statutory declaration by the complainant that the above is correct to the best of their knowledge.

- an internet intermediary in receipt of a valid mandatory concerns notice may make an offer to make amends to the complainant (but is not required to);

- if the internet intermediary makes an offer to make amends:

  - the offer *must* include an offer to notify the originator to:
    - seek consent to provide identifying information of the originator (to the extent that it exists and is identifiable); or
    - give them an opportunity to remove the content-at-issue;

  - the offer *may* include any other offer that the intermediary deems appropriate, including offers to remove content, consent to a preliminary

discovery order, and any of the mandatory and voluntary matters in the legislation (each of which would become voluntary).[39]

- It is a defence to an action for defamation against the internet intermediary if the complainant accepts the offer to make amends (and the offer is then carried out), or the complainant fails to accept a reasonable offer to make amends.

Given the Stage 1 amendments to the MDPs have made the concerns notice mandatory prior to commencing litigation (unless the complainant obtains leave), our proposed approach would prevent a complainant from commencing proceedings against the internet intermediary unless they have either obtained judgment against the originator, or cannot identify the originator (as the complainant cannot otherwise issue a concerns notice to an intermediary).

The advantage of our recommendation is that it would preserve the existing procedures for originators, and creates a complementary process for intermediaries directed towards achieving the primary objectives associated with complaints to intermediaries under the complaints notice process: identification of the originator, and/or removal of the content-at-issue. This is achieved without introducing new concepts, and with minimal amendments to the current system.

Issue 4: Power of courts to order that material be removed

We do not support the expansion of power to courts to take down orders against internet intermediaries that are not parties to defamation proceedings. If an Australian court declares content is unlawfully defamatory, Facebook ensures that content is restricted or geo-blocked from our services in Australia - so it is not necessary to ensure that we are taking action to restrict access to the content on our services.

We would be concerned at any suggestion that courts should be enabled to compel blocking of material elsewhere in the world, outside of Australia's jurisdiction. Australian legal requirements for empowering courts to order removal of online material within Australia (e.g. injunctions) are well established and no further changes are required.[40] While the legal standard for obtaining an injunction is arguably higher in

---

[39] For example, an offer to: (1) publish or join in publishing a reasonable correction, clarification or additional information about the matter in question (clause 15(1)(d) MDAP); (2) pay the expenses reasonably incurred by the aggrieved person (clause 15(1)(f) MDAP); (3) publish or join in publishing an apology (clause 15(1A)(a) MDAP); (4) pay compensation for economic or non-economic loss (clause 15(1A)(b) MDAP), etc.

[40] Generally, that (1) there is a serious question to be tried, or that the plaintiff has made out a prima facie case; (2) the plaintiff will suffer irreparable injury for which damages will not be adequate compensation

defamation actions (given the public interest in free speech),[41] this is not unreasonable in the circumstances.

“Harm” under Australian defamation laws occurs in the place where the damage to reputation occurs (ie within Australia). To order worldwide blocking would be to apply Australian legal standards outside Australia without justification, including without consideration of other legal norms that operate outside Australia. Even if courts are granted this power, other jurisdictions may (rightly) not recognise it. For example, a court in the United States would be unlikely to give effect to that order without being satisfied that the first amendment was adequately protected.

Issue 5: Power of courts to order that internet intermediaries reveal the identity of originators

We do not believe changes governing when a court may order that an internet intermediary disclose the identity of a user are required.

The procedural rules for obtaining discovery orders are well established, and "*Norwich Pharmacal*" orders are available in Australia in suitable cases.[42]

In relation to seeking orders against entities outside Australia, the legal standard complies with international legal principles. Legal processes should be issued to foreign entities in compliance with principles of international comity, and only in exceptional circumstances. Relevantly, "exceptional circumstances" arguably include where the legal process is limited to production of basic identification material.[43] Thus there is a well-established process for obtaining identifying information about a proposed defendant, which complies with international legal principles.[44]

Any jurisdictional issues arising in obtaining such orders can be overcome by commencing proceedings in the appropriate jurisdiction (e.g. the Federal Court). There may also be scope to simplify the preliminary discovery process through

---

unless the injunction is granted; and (3) the balance of convenience favours granting the injunction. See *Australian Broadcasting Corporation v O'Neill* (2006) 227 CLR 57; and *ABC v Lenah Game Meats* (2001) 208 CLR 199.

[41] *National Mutual Life Association of Australasia Ltd v GTV Corpn Pty Ltd* [1989] VR 747; *Chappell v TCN Channel Nine Pty Ltd* (1988) 14 NSWLR 153; *Clarke v Queensland Newspapers Pty Ltd* [2000] 1 Qd R 233; *Jakudo Pty Ltd v South Australian Telecasters Ltd* (1997) 69 SASR 440, [442]–[443].

[42] *Bentley Fragrances Pty Ltd v GDR Consultants Pty Ltd* (1985) 11 FCR 29, 34 – 35.

[43] *Ceramic Fuel Cells Limited (In Liq) v McGraw-Hill Financial Inc* [2016] 401; *Ceramic Fuel Cells Ltd (in liq) v McGraw-Hill Financial Inc* [No 2] (2016) 245 FCR 362.

[44] *Kabbabe v Google LLC* [2020] FCA 126.

amendments to procedural rules rather than through the vehicle of reforms to the MDPs.

Other issues

Ideally, the liability of internet intermediaries (or defences available to intermediaries) under Australian law should align with those in major foreign jurisdictions such as the US and the EU. Policymakers should consider the position in the EU, which exempts hosting platforms from liability where they do not have actual knowledge of illegal activity or information and, upon obtaining such knowledge, act expeditiously to remove or to disable access to the information.