

Social Media and Defamation Law

Contents

What is Social Media?	2
How has the law of defamation dealt with other media?	3
The law's response to new technology	6
What is different about social media platforms?.....	6
The arrival of the internet	7
Regulation of new technologies.....	8
Regulation of social media.....	10
Voller's case	10
<i>Google LLC v Defteros</i>	12
The issue of publication is fact specific	13
The public and policy reaction to <i>Voller</i>	14
Whose domain is it?.....	15
The separate issues of publication and innocent dissemination.....	15
Problems besetting the concept of publication on social media.....	16
Reforms to Australia's defamation laws	16
Other legislative developments.....	20
The federal government's plans to address online misinformation	21
An international perspective	21
The architecture of local laws about internet intermediaries' liability for third-party content	22
How are internet intermediaries different?	23
The limits on legislation.....	25
Encouraging moderation.....	26
Potential civil law liability as an incentive to moderate	27
An enduring problem - to censor or not to censor?.....	28
Take down orders	29
Defamatory meaning.....	30
Search results and snippets	32
Hyperlinks.....	33
Qualified privilege and public interest defences.....	33
Litigation and social media.....	35
The difficult question of who is the publisher?.....	37
Encouraging moderation - The USA v The rest of the world	37
A right-wing challenge to self-moderation	39
Conclusion	41

Social Media and Defamation Law

The Hon Justice Peter Applegarth AM¹

This is a big topic for a short talk. It raises these issues:

- What is social media and how is it different to other kinds of media?
- How should existing legal principles apply or adapt to it?
- What new laws are needed?

What is Social Media?

How do we define social media, given its diversity? Social media ranges from the lonely, poor blogger who creates content to multi-billion dollar social media platforms that host and communicate content created by others.

The Oxford English Dictionary defines social media as “websites and applications that enable users to create and share content or to participate in social networking.”² Social media includes a multiplicity of actors and therefore a multiplicity of sins. Some actors create content and communicate it on blogs or their personal websites. Other actors host and communicate content that is created by users.

Social media started with ‘weblogging’ communities. Disputes over offensive content and whether it should be regulated at all started at the infancy of the internet.³ Today, blogs are ubiquitous. Sites like Reddit essentially aggregate bulletin boards that are overseen by volunteer moderators who apply rules for that group. Platforms include social networks like Meta’s Facebook; hosted communities (e.g. when an online newspaper creates a site for readers’ comments and uses a technology like Facebook to do so); collaborative networks like Wikipedia; content communities like YouTube that allow users to share a video and rate or comment on content, Instagram and TikTok; and many others.⁴ Joseph Collins notes that increasingly the distinction between these forms of platforms are breaking down as platforms expand their features, for instance, with messaging services being incorporated into networks (e.g. WhatsApp) or expanding to become social media networks (e.g. Snapchat).⁵

Typical features are apparently free access (“free access” has a privacy price as users’ information is monetised and sold to advertisers and others); platforms’ dependence on user-generated content; sharing of content with the assumption that anything that is posted is available to the world, unless the person specifies otherwise in their privacy settings; and non-negotiable contractual terms that authorise use, describe the platform as a content host, indemnify it, and seek to ensure ownership and responsibility for posting reside with the user who posts content.⁶

¹ Judge, Supreme Court of Queensland. A paper presented on 17 August 2023 to *Current Legal Issues Seminar 2: Social Media and Defamation Law*, hosted by the Bar Association of Queensland, the University of Queensland and the Supreme Court Library Queensland.

² *Oxford English Dictionary* (online at 2 August 2023) ‘social media’.

³ J. Ronson, ‘A Scottish Jewish Joke’, *Things Fell Apart*, BBC Radio (7 December 2021) <https://www.bbc.co.uk/sounds/play/m00127xv>.

⁴ J. Collins, ‘Social Media and the Law’ in P. George (ed), *Social Media and the Law* (Lexis Nexis, 2023) at [1.24]-[1.28].

⁵ *Ibid* at [1.26]-[1.27].

⁶ *Ibid* at [1.29]-[1.40].

Let us focus on social network platforms that host third-party content, rather than social media content creators. How should we perceive them and how should the law treat them? Judges, like other human beings, like to categorise things. We tend to treat a new technology as analogous to a familiar technology.

Is a reader's comment posted to the Facebook page of a newspaper in a different category to a letter to the editor in its hard copy edition?

Is X (the platform formerly known as Twitter) analogous to a village noticeboard, except the village is a planet and the owner of the noticeboard profits from what is published?

Should social media platforms that host content created by a third party be placed in the same category as other media entities that can moderate or remove content that is supplied by third parties?

Should their moderation policies and practices be self-governing, subject possibly to transparency requirements so we know what they are? Should entities be regulated by an agency that oversees Codes of Conduct about moderation and removal of harmful third-party content? Does the law of defamation have any role to play in creating incentives to moderate content and liability if a content host declines to moderate or remove harmful content after being on notice of it?

Is social media simply a different form of communications technology that should be governed by the same general principles and laws that apply to older forms of technology with, if necessary, certain statutory immunities or new defences that recognise their differences?

How has the law of defamation dealt with other media?

Defamation is a tort of strict liability. The actionable wrong is the publication. A publisher's liability does not depend on their knowledge of the defamatory content, an intention to defame or fault.⁷ Anyone who intentionally assists in the process of communicating the matter that contains defamatory content is liable, regardless of whether the person knows that the matter contains that content.⁸ Expressed differently, anyone who voluntarily participates in the communication of the publication that is read or understood by a third party is *prima facie* liable.

The strictness of liability for persons who do not create the content and who do not know of it led to the development of the common law defence of innocent dissemination and the enactment of innocent dissemination defences for certain subordinate or subsidiary publishers. The defence applies to distributors, booksellers, printers and others. In essence, the common law defence applies if such a publisher did not know that the publication was of a character that was likely to contain defamatory material and their lack of knowledge was not due to negligence on their part. The defence exists to protect parties who have no practical ability to control or supervise the material that they publish.⁹

Still, not everyone who facilitates the publication of defamatory matter is taken to be a publisher. It is not sufficient that a defendant merely plays some passive instrumental role in the process of publication.¹⁰

⁷ *Fairfax Media Publications Pty Ltd v Voller* (2021) 273 CLR 346; [2021] HCA 27 at [27] (“*Voller*”).

⁸ *Voller* at [68].

⁹ *Defteros v Google LLC* [2021] VSCA 167 at [144].

¹⁰ *Voller* at [18].

Some communicators are not treated as publishers, and others like printers and distributors who are publishers might rely on a defence of innocent dissemination. The postal worker who delivers a defamatory letter or the telephone company that sends a defamatory statement down its wires facilitates the publication of defamatory matter and, in a sense, is a publisher. The law, however, never saw it that way. To hold the humble postman, the former Postmaster General, or Australia Post to be a publisher and liable for defamatory letters would be an affront to common sense. The law treats them as mere conduits. To treat them and Telcos as publishers and require them to have systems to monitor the contents of letters and phone calls in order to engage an innocent dissemination defence would have made their businesses unworkable. It also would have been an unjustified intrusion on the privacy of communications.

The development of the defence of innocent dissemination recognised that it is unreasonable for some persons who do not reasonably know of the defamatory statements that they publish to be liable under a tort of strict liability. The argument was that one should not “shoot the messenger”, at least when the messenger or facilitator did not reasonably know of the defamatory content of the publications they facilitated.

The same impulse that protects certain printers, distributors and other people who communicate another party’s content may incline a legal system to view modern disseminators who operate social media platforms as not being publishers or not being publishers who should be liable for content that they did not create and of which they are not aware.¹¹

Another basis for liability in defamation arises from the noticeboard case of *Byrne v Deane*.¹² It rests on a failure to remove allegedly defamatory content posted by another. The directors of the golf club in that case were found to be publishers for having failed to remove the material when they had power to remove it. They had not participated in the primary act of publication, but their conduct or inaction after becoming aware of it meant that they could be said to have participated in a continuing publication.¹³

We have had noticeboards on the walls of halls, clubs and colleges for ages. The general law developed to define when an entity that controls a physical noticeboard is liable for something that is posted on it. In the last few decades, we have seen the development of billions of online noticeboards: starting with rudimentary online discussion boards, chat rooms, then Facebook pages, then tweets. In *Tamiz v Google Inc*,¹⁴ the host of a blogging site was held to be (arguably) a publisher *after* it received notice of the defamation, based on *Byrne v Deane*. However, it has been said that noticeboard cases like *Byrne v Deane* are not useful to explain the involvement of others in publications in very different circumstances such as internet platform providers and hosts of Facebook pages.¹⁵ There is a critical difference between cases involving parties who have played *no* role in facilitating publication of defamatory matter prior

¹¹ Early case law in the UK treated an internet service provider that provides its subscribers with access to discussion forums in the same legal category as distributor: *Godfrey v Demon Internet Ltd* [2001] QB 201 at 208-209. Legislation subsequently addressed the status of an ISP that merely provides customers with access to the internet as conduits governed by *Electronic Commerce (EU Directive) Regulations, 2002*, reg 17. In Australia, such an internet service provider that does no more than carry content is unlikely to be treated as a publisher under the common law: *Google Inc v Duffy* [2017] SASCFC 130; (2017) 129 SASR 304 at [138]-[139]. This will be confirmed by proposed laws that will provide a conditional, statutory exception from defamation liability for mere conduits (including ISPs), caching and storage services.

¹² [1937] 1 KB 818.

¹³ *Voller* at [50]-[54]

¹⁴ [2013] EWCA Civ 68.

¹⁵ *Voller* at [54] citing *Oriental Press Group Ltd v Fevaworks Solutions Ltd* (2013) 16 HKCFAR 366 (internet platform providers) and *Murray v Wishart* [2014] 3 NZLR 722 (hosts of a Facebook page).

to becoming aware of it and cases where a party facilitates third parties publishing material on the party's site.¹⁶

If social media platforms are seen as the functional equivalent of an entity that controls material that is posted on a physical noticeboard, then the volume of postings and the inability to monitor all of them may support the creation of immunities from liability, at least until the operator of the platform is placed on notice of the defamatory content of a posting. But what should the platform operator's liability, if any, be once it receives a complaint about a defamatory posting and knows that it is hosting and communicating harmful defamatory content?

Exposure to liability may encourage the platform to remove the post at the expense of freedom of expression, including posts that are true and about a matter of legitimate public interest.

On the other hand, insulating platforms from legal liability and leaving moderation and removal to self-regulation equips malicious keyboard warriors with a platform to defame. Without the platform there would not be the mass communication and the harm.

The plea "Don't shoot me, I'm only the messenger" may work for postal workers. But should it work for a social media platform that provides a platform that can inflict massive harm on individuals who are defamed, knows of the defamatory content, and refuses to remove it. These platforms adopted a business model that has a huge competitive advantage over legacy media that moderate content. They have killed a large part of traditional media. They profit by hosting controversial content. More controversy, more clicks, more money for the operator of the platform.

Having lost the *Voller* case for encouraging readers to post comments on the Facebook pages that hosted their newspaper articles and broadcasts, legacy media want to play by the same legal rules as Facebook and other content hosts. Legacy media want their social media pages to be treated like those of an internet intermediary that hosts a third party's content. They want a reader's comment on their Facebook page to be treated differently to an old-fashioned letter to the editor.

As for social media platforms, the current issue is whether they:

- are left to regulate themselves, possibly with new laws that mandate disclosure of policies and practices to remove disinformation and misinformation;
- are regulated in some other fashion;
- are exposed to civil liability for defamation if they decline after a reasonable period to moderate or remove content following a complaint.

I will return to pending legislation in Australia that has been approved by our Attorneys-General about the liability of internet intermediaries for third-party content. First, it is worth thinking more broadly about the law's response to new technology and different approaches to regulation. I also want to map the path that led to the mess that the law now is in, including the passage of protective federal laws for operators on the internet. They were enacted at a time when no one had heard of Facebook or Twitter because they had yet to be invented.

¹⁶ *Voller* at [103].

The law's response to new technology

The application of defamation law to social media is simply one example of the law responding to new technology: something the law has done for centuries. For instance, negligence law adapted to the advent of the railways. The general law and regulatory law developed to deal with modern pharmaceuticals.

Defamation law responded to the power and prevalence of the printed word (with the distinction between libel and slander), the arrival of mass circulation newspapers, and developments in printing and broadcasting. Television and radio had to fit into the law. Oral defamations over the airwaves that ordinarily would be slander were deemed by statute to be libel. Courts appreciated, however, the transitory nature of broadcast communications compared to the printed word.¹⁷ This bears on the defamatory meanings that an ordinary listener or viewer would draw. Contemporary judges have had to consider how users of social media understand the messages they receive and the defamatory meanings they might reasonably derive from the relevant medium.¹⁸

With traditional media, technological developments in printing saw the end of the era of printers setting every letter in hot metal (when printers were taken to know what they were printing). Automated printing meant printers had no reasonable opportunity to read in advance what was speeding off the presses once an editor ordered “print”. The defence of innocent dissemination developed to protect printers who took reasonable steps to know what they were printing. That defence also assisted distributors who did not have reasonable grounds to know that what they were distributing was defamatory.

Broadly speaking, and taking account of innocent dissemination defences, defamation law, both at common law and under statute, makes a broad distinction between:

- the creators and controllers of content, so-called “originators” like journalists, editors, and the owners and operators of legacy media like newspapers and broadcasters; and
- those who host or communicate someone else’s content, and who do not have the capacity to exercise editorial control over the content, including some communicators like Telcos who are not treated as publishers.

The innocent dissemination defences that are available to the second category are not immunities from liability, as are conferred by certain statutes that provide “safe harbours”. They are defences that require a subsidiary or subordinate publisher to prove certain things. They rest on the idea that in some circumstances a person should not be liable for unknowingly communicating defamatory matter, but ignorance of the content is no defence where it was reasonable to monitor the content so as to be able to moderate or remove it.

What is different about social media platforms?

Why should social media platforms that host and communicate defamatory content be in a different position to a bookshop, a library, or a golf club that encourages its members to use its noticeboards? Should they only have an innocent dissemination defence if they have reasonable systems that monitor what they publish?

¹⁷ *Amalgamated Television Services v Marsden* (1998) 43 NSWLR 158, 165-166.

¹⁸ See section on Defamatory meaning at pp 30-32 below.

One argument for treating social media platforms differently is that the task of monitoring and moderating content is beyond them and the economic and social benefit that they generate by facilitating social networking and hosting third-party content makes it unreasonable to expose them to liability for a failure to monitor and remove.

Ultimately, this is the type of argument that saw certain social media entities given a broad statutory immunity in the US, and conditional protections under federal law in Australia and under laws in the UK and other countries.

The arrival of the internet

Australia and other countries addressed the arrival of the internet by enacting laws that gave internet content hosts varying degrees of protection. In the US, s 230 of the *Communications Decency Act* 1996 gave an effective immunity to social media platforms from liability. In Australia, clause 91 of Schedule 5 of the *Broadcasting Services Act*, enacted in 1999, overrode the general law and any State or Territory statute to the extent to which such laws imposed liability for illegal content where the host or service provider was unaware of the nature of the content. It overrode any common law rule or State or Territory law that requires an internet content host to monitor or make inquiries about the online content that it hosts. It aimed to protect entities that hosted someone else's website on their server, discussion groups, such as community or professional groups with an interest in a certain topic, and internet service providers like OzEmail.

When that law was enacted Mark Zuckerberg was aged 15 and still not at Harvard. No one in Australia had a Facebook friend. Twitter did not exist. Those of us who dialled onto the internet would select our search engine like AltaVista or Yahoo. This was before Google won the arms race to be the dominant search engine.

The inclination of legislators 25 years ago to protect parties that provided internet services is understandable. It made sense to provide that kind of protection to advance the use of the internet for the greater public good.

Then, as now, a few participants monopolised print media and there were high barriers facing participants who wanted to start new newspapers. Dissenting or diverse voices could not easily get their message to the public via Mass Murdoch. The internet was seen as a means for these people to get their message to the world and compete in a new marketplace of ideas. One felt like Wordsworth did at the start of the French Revolution, "Bliss was it in that dawn to be alive."

The popular imagination and the prevailing public policy were that the internet should be a free electronic frontier. Exposing internet content hosts to liability seemed like a bad idea at the time. To expose them to liability for defamation seemed like "shooting the messenger", something the law did not do to postal workers, librarians, newsagents, or telephone companies.

At the time, legislators did not have Facebook or Twitter in mind. Internet content hosts that ran commentary bulletin boards for community or professional groups provided a valuable service and did not need to be licensed or regulated in the way that some industries are.

The print media is not licensed due to an historic aversion, dating back to Milton, to the evils of state control over content. The kind of licensing of television and radio stations that was developed in the 20th Century is obviously inappropriate for the number and diversity of social media participants.

The historic, principled opposition to licensing newspapers on free speech grounds carried over to social media participants, big and small. Also, there is the practical matter that some social media platforms are global entities and may not be influenced by a domestic law that licenses or otherwise purports to regulate them. They claim to have no legal presence in that jurisdiction and any fines or penalties for regulatory breaches are ignored because they, their servers and their money are outside that jurisdiction. Also, American laws effectively make foreign judgments unenforceable against the media.

The last decade has forced communities and governments to think about the regulation of social media platforms and their statutory immunities. The concern has been prompted by hate speech, revenge porn, industrial-scale invasions of privacy, the real time hosting of terrorist crimes as occurred in Christchurch, and conspiracy theorists who make wild claims, like the ones who defamed the parents of children who were shot at the Sandy Hook Elementary School. The concern has been heightened by the prevalence of organised misinformation and disinformation on a scale that threatens democratic institutions.

The law, as usual, is addressing a technology that is evolving, creating social benefits and creating social and individual costs. The law lags behind technology.

Regulation of new technologies

New technologies bring benefits and usually inflict costs. Railways brought great social and economic benefits, along with dead cattle, burnt wheat fields, and injured passengers. Railways were regulated and exposed to civil liability under the general law. The fact that an industry does more social good than harm should not insulate it from regulation. Pharmaceuticals do more good than harm, but the pharmaceutical industry is regulated. Professional advisers about financial products are regulated and licensed, subject to consumer protection laws and exposed to civil liability under the general law.

The advent of Artificial Intelligence, including the current panic about ChatGPT, has seen calls for its regulation. Professor Noah Feldman recently wrote about that issue and outlined the regulatory possibilities to deal with it.¹⁹ Possible forms of regulation for a new technology include:

- Prohibition and intense regulation for a technology that poses an existential threat to humanity, like nuclear weapons;
- Criminal laws that punish people who use technology to commit fraud (hence the term “wire fraud”) or use telecommunications or the internet to send material that harasses and seriously harms individuals;
- Statutory civil regulation with violations punishable by fines or a pecuniary penalty;
- Administrative rules that are enforced by a regulatory agency; and
- Civil liability under a tort system that may require the user to exercise reasonable care, or impose some other standard of civil liability, subject to defences and immunities that are either devised by judges or enacted by legislators.

¹⁹ N. Feldman, ‘Regulating AI will be essential. And complicated’, *The Business Standard* (4 April 2023) <https://www.tbsnews.net/features/panorama/regulating-ai-will-be-essential-and-complicated-610662>.

There is a regulatory smorgasbord. Some activities have regulators. Product liability laws can be enacted that relieve the injured from proving negligence and make industry participants pay for the harms their products cause. In some situations, we rely on the general law to do the heavy lifting and, if the need arises, smooth its rough edges by enacting statutory modifications to common law duties or by creating new statutory defences. Often there is a combination of regulation through regulatory agencies, statutory duties, the general law and non-enforceable codes of practice.

In some circumstances, an activity is thought to be so beneficial to society that its existence and growth are protected by laws that in certain circumstances immunise it from liability under the general law or obligations imposed by statute. Imagine, however, that in the 19th Century legislators had been so entranced by the social and economic benefits of railways that they immunised railway companies from any liability, including for the cattle they killed, the citizens they ran over or the passengers they carelessly killed.

Yet, a similar broad immunity was given by the US Congress to social media platforms and other internet intermediaries in 1996. Other countries, including Australia, passed less protective laws to encourage the economic and social benefits of the internet. Twenty-five years later we are trying to work out what they mean and to which content hosts they apply.

As noted, the absence of regulation of social media platforms and the statutory immunities from liability that some of them enjoy were the result of a social consensus that the internet brought a new era for freedom of expression and that its potential for enhancing communications between its users justified shielding its operators from civil liability.

Section 230 of the US Act states:

“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.”

Those are said to be “The Twenty-Six Words That Created the Internet”.²⁰ Section 230 did not create the internet, but it did foster a legal environment and economic model for modern social media platforms. They rely on users’ content. They monetise a user’s personal information by algorithms that target ads and sell access to their users as targets.

Over time, s 230 of the US Act has been treated as a grand charter that social media operators should be immune from the threat of liability. Yet, its purpose was more limited. The statute was not named “The Internet shall be Free Act”. Its name gives the game away: the *Communications Decency Act*. It was about pornography. Sadly, one cannot talk about the early days of the internet without talking about pornography. Those who sponsored s 230 now say that they had a vision of it applying to the social media platforms that have developed since its enactment. The better view is that it was less ambitious.²¹

²⁰ J. Kossuff, *The Twenty-Six Words That Created the Internet* (Cornell University Press, 2019).

²¹ D. Citron, ‘How To Fix Section 230’ (2023) *Boston University Law Review* (forthcoming) https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4054906; An informative interview with Professor Citron can be read or heard on the Amicus Podcast ‘SCOTUS on the Internet: It’s Complicated’, SLATE Podcasts (25 February 2023) <https://slate.com/podcasts/amicus/2023/02/twitter-and-google-at-the-supreme-court-left-most-of-the-justices-scratching-their-heads>.

Regulation of social media

My topic is social media and the law of defamation. However, we need to address defamation on social media as part of a broader discussion about how regulators and courts address other forms of harm on social media: hate speech, fake news, sensitive private information that is not defamatory, and statements that are not unlawful, but simply harmful to vulnerable individuals like victims of sexual abuse and children.

The arrival of the internet raised policy issues about the regulation of new forms of communication on the internet, including the extent to which they should be regulated by an authority that regulated other forms of communication. One policy challenge was competition law, leading to the ACCC's Digital Platforms Inquiries. Another was privacy, including the regulation of data privacy. Another was the posting of naked images of former sexual partners as a form of revenge. Another was the hosting and communication of material that incites racial hatred.

The risk of liability, and therefore the perceived need for regulation, arose if a content host or internet service provider took the responsible step of attempting to monitor and moderate content. That exercise of control over content placed it in legal jeopardy of being deemed to be a publisher. Therefore, laws were enacted to avoid that risk.

Laws enacted to protect innocent internet content hosts from liability for hosting certain offensive material, particularly pornography, created a safe harbour for a variety of internet content hosts against a broad range of potential liabilities, including liability for defamation under the common law and under State and Territory statutes.

The deepest safe harbour was that created by the US Congress in 1996. The Australian safe harbour created in 1999 by the passage of Schedule 5 to the *Broadcasting Services Act* (Cth) was a shallower safe harbour of uncertain depth. It has been relocated to the *Online Safety Act* 2021 (Cth). Confusion remains over who that federal law protects and the extent to which it shields certain internet content hosts from liability in publishing third-party content.

Clarification of the liability of internet intermediaries (not just "internet content hosts" within the meaning of the 1999 law) for third-party content has been on the agenda for a long time. It led to the release in August 2022 of proposals by the Attorneys-General for new laws. These are known as the Stage 2 Part A Reforms. Subject to final agreement, new provisions are intended to commence from 1 January 2024. In the meantime, Australian courts have had to grapple with issues of who is a publisher. This culminated in the High Court's decision in September 2021 in *Voller*, and its decision in August 2022 in *Google LLC v Defteros*,²² about whether Google was a publisher of certain search results.

Voller's case

Mr Voller sued several media entities. Each entity operated a Facebook page on which it posted news content. The media entities had posted news stories about Mr Voller and invited comments from Facebook users. Third parties posted defamatory comments. Mr Voller claimed that the media entities were liable as publishers of the defamatory comments. The primary judge held that the media entities were publishers of the comments. This was confirmed by the New South Wales Court of Appeal.

²² [2022] HCA 27 ("*Defteros*").

The High Court also found that media entities that create and maintain Facebook pages are liable as publishers for defamatory comments posted on them by third-party users. Kiefel CJ, Keane and Gleeson JJ found that the entities were publishers because they had facilitated, encouraged and thereby assisted the posting of comments by users. They rejected the contention that publication is dependent on knowledge or intention.²³

Gageler and Gordon JJ likewise concluded that the media entities were publishers, noting that “the appellants’ attempt to portray themselves as passive and unwitting victims of Facebook’s functionality has an air of unreality”.²⁴ The media defendants encouraged user engagement and derived commercial benefit from Facebook’s functionality which facilitates this interaction. As a result, they had to bear the legal consequences of such activity.²⁵

Edelman J concluded that the defendants were not publishers of the “uninvited words written on their Facebook pages”.²⁶ They had not invited or requested the comments and had not manifested an intention to publish them. According to Edelman J, intention to publish “is concerned with an intention to act rather than with any of the consequences of the act”.²⁷ Although the entities had not published the comments, it was relevant to consider whether they had assisted users with a common intention to publish the comments. Edelman J concluded that “merely creating a page and posting a story with an invitation to comment” was insufficient to manifest a common intention to publish comments that were unrelated to the posted story.²⁸

Steward J found that the media entities had facilitated the publication of the comments by creating Facebook pages and making their own posts, but that such acts did not make the entities publishers of all third-party comments made on their pages.²⁹ A different situation would arise if the content of the posts was likely to procure, provoke or conduce a defamatory response by third-party comments. This would be a question of fact and if such a finding was justified by the content of the posts, the media entities would be taken to have participated in publication.³⁰

Importantly, *Voller* concerned an appeal to the New South Wales Court of Appeal, and then a subsequent appeal to the High Court on the threshold and preliminary issue of whether the media defendants were publishers. Those courts did not consider whether the defendants were able to establish an innocent dissemination defence. The primary judge was criticised for having addressed a later question about whether each media defendant was a primary publisher or what is variously described as a subsidiary, subordinate or secondary publisher. That question had relevance to a potential innocent dissemination defence. Rothman J opined that they were primary publishers. The High Court did not address that issue or consider, other than in passing, the Commonwealth safe harbour provision that started life as clause 91 of Schedule 5 in 1999.

In the New South Wales Court of Appeal Basten JA considered that provision in surveying the legal landscape. Several non-party media companies sought leave to appear in the proceedings and raised clause 91. Basten JA noted that clause 91 requires that, to the extent of any inconsistency, Commonwealth law will be paramount.³¹ Therefore, any State law which imposes liability for defamation in the circumstances covered by clause 91 will not apply.³²

²³ *Voller* at [27].

²⁴ *Voller* at [102].

²⁵ *Voller* at [102].

²⁶ *Voller* at [109].

²⁷ *Voller* at [115].

²⁸ *Voller* at [139], [142].

²⁹ *Voller* at [177], [179].

³⁰ *Voller* at [183].

³¹ [2020] NSWCA 102 at [16]; (2020) 105 NSWLR 83.

³² At [17].

However, clause 91 only has that effect where the internet content host is unaware of the nature of the content. Consequently, if the content host is aware of the nature of the content, then the law of the State or Territory will continue to apply.³³

Basten JA also highlighted the fact that whether a State law or the general law ceases to have effect by operation of clause 91 depends on the party being an “internet content host” with respect to the third-party posts.³⁴ The definition of “internet content host” required the content to be hosted “in Australia” and this was ambiguous. On one view, it required the data to be stored on a server in Australia. An alternative was that it deals with content hosts that have a presence in Australia and are amenable to local laws.³⁵ On this preferred view, an overseas media company that operates from the US is not an “internet content host” for the purposes of clause 91.³⁶

Google LLC v Defteros

Long-settled definitions of publication and who is a publisher may produce different conclusions among reasonable minds. After *Voller* settled the test for publication, the High Court revisited the issue of publication in *Google LLC v Defteros*.

George Defteros sued Google after it refused to take down a hyperlink to an article published by *The Age*. The article referred to charges that were brought against Mr Defteros but subsequently withdrawn.³⁷ At trial and on appeal, the item was held to be defamatory and Google liable as a publisher of it.³⁸ On a further appeal to the High Court, Google argued that the Victorian Court of Appeal erred in concluding that it was a publisher of the content and that the defences of common law and statutory qualified privilege and innocent dissemination were not made out.³⁹

The majority found that Google was not the publisher of defamatory material which is merely hyperlinked in organic search results.⁴⁰ Kiefel CJ and Gleeson J noted that a “search result is fundamentally a reference to something, somewhere else. Facilitating a person’s access to the contents of another’s webpage is not participating in the bilateral process of communicating its contents to that person.”⁴¹ The majority referred to the decision of the Canadian Supreme Court in *Crookes v Newton*⁴² which held that a hyperlink is merely a reference to the linked content and does not itself publish the content.⁴³

The majority rejected the Court of Appeal’s suggestion that by incorporating the link into a search result that contained a snippet of an article “Underworld loses valued friend at court”, Google effectively enticed users to click on the hyperlink and access the defamatory matter.

³³ At [29]-[30].

³⁴ At [17].

³⁵ At [22]-[23].

³⁶ At [29].

³⁷ *Defteros* at [1]-[3].

³⁸ *Defteros v Google LLC* [2021] VSCA 167.

³⁹ *Defteros* at [8].

⁴⁰ These are unpaid search results that are found and ranked algorithmically based on relevance, uninfluenced by advertisers.

⁴¹ *Defteros* at [53].

⁴² [2011] 3 SCR 269.

⁴³ *Defteros* at [42].

Gagelar J agreed that merely providing a search result that includes a hyperlink does not “direct, entice or encourage the searcher to click on the hyperlink”.⁴⁴ There may, however, be circumstances where a series of search results amount to such an enticement. Also, Gagelar J considered that a sponsored link placed Google in the position of publishing another’s advertisement.

Edelman and Steward JJ joined the majority finding that there was no actual communication, authorisation or ratification of the defamatory article by publishing the hyperlink.

Keane J dissented, noting that Google participates in the process of publication by providing near-instantaneous access to linked content. Keane J noted the broad application of the publication rule has long been a feature of Australian defamation law. Google’s participation in the provision of linked content was sufficient to come under the broad umbrella of ‘publication’.⁴⁵ Google’s search results directed its users to *The Age’s* article through the hyperlinks it provided. This direct and near instantaneous access occurred as Google intended by the operation of its search engine. Google thereby participated in the publication of *The Age’s* article.

Gordon J reached the same dissenting conclusion as Keane J about Google’s intention to facilitate access to news articles and its participation in the publication. Gordon J stated:⁴⁶

“Having taken action to obtain a commercial benefit by creating and operating a search engine that facilitates access to news articles, it cannot deny that it is involved in the publication of those new articles.”

The High Court’s 5:2 decision in *Defteros* about hyperlinks aligns Australia with other jurisdictions including the UK and Canada. However, the issue before the Court was limited to organic search results where the hyperlink is clicked on by the reader to access defamatory content. It did not consider liability for defamatory content in snippets within the text of the search result itself or content contained in sponsored search results. The decision leaves open the possibility that Google may be liable as a publisher of sponsored hyperlinks that result from an advertising agreement between the search engine and the content creator.⁴⁷

The liability of search engines like Google for “organic search results” is to be addressed by Stage 2A Recommendation 2, noted earlier. No issue of moderating content arises in the case of “organic search results” because the search engine service is unable to modify or remove the things it locates.

The issue of publication is fact specific

An unhelpful question is to ask, “Is Google liable for defamatory matter that it helps communicate?” Its potential liability depends on:

- the acts or omissions that are alleged to constitute publication by it or its participation in a third party’s publication;

⁴⁴ *Defteros* at [74].

⁴⁵ *Defteros* at [103]-[104].

⁴⁶ *Defteros* at [112].

⁴⁷ *Defteros* at [69]-[70].

- the time of alleged publication, particularly whether the publication is alleged to have occurred before it received a complaint or otherwise became aware of the content;
- what it communicated, for instance, whether the complaint is that:
 - its search engine results provided a hyperlink by which defamatory matter may be found by clicking on it;
 - its search engine produces an autofill suggestion in a box after typing the claimant's name or produces a page of results with snippets of information;
 - it is hosting a blog or discussion group;
 - it is hosting a defamatory video on its YouTube platform, as it did when it hosted the FriendlyJordies videos that seriously defamed Mr Barilaro.⁴⁸

Authoritative decisions like the 5:2 ruling in *Google LLC v Defteros* need to be carefully understood since they relate to only certain kinds of search engine results.

If a party is, however, a publisher of certain search engine results, the tricky question is what do those results mean, if anything, to an ordinary reader? Are they *capable* of conveying the *defamatory meaning* contended for by the claimant (a question of law) and do they in fact do so (a question for trial and, therefore, for a jury if there is one). Those questions involve issues about how ordinary, reasonable users of social media platforms interpret what they read, including the attention and analysis they devote to them. I address them later.

The public and policy reaction to *Voller*

The appellants in *Voller* were media defendants with deep pockets. However, the issues that *Voller* raised about liability for third-party content that is posted on a publicly accessible Facebook page had obvious implications for small publishers, “citizen journalists”, community organisations and many individuals.

In reacting to their defeat in the High Court, media organisations and their allies expressed surprise that someone should be liable for a third party posting defamatory statements on a Facebook page upon which the defendant had posted its news articles or online editions of its newspaper.

Some of us whose early careers included defending newspapers that printed letters to the editor tended to see the issue differently. We did not see the media organisation's *prima facie* liability as categorically different to the liability of a newspaper that publishes a letter to its editor or a talk-back radio station that allows telephone callers onto its airwaves. We search for analogies with familiar things. Meagher JA and Simpson AJA did so when they observed in *Voller*:

“... the third parties' composition, comment and publication occur without any further and specific acts of participation by the applicants. That is because the facility subscribed for and used by the applicants enables all of that to occur without their further intervention. In those respects that facility is similar to one offered by a talkback radio station broadcasting live commentary from listeners on the telephone.”

⁴⁸ *Barilaro v Google LLC* [2022] FCA 650.

Whose domain is it?

There are two competing perspectives on the issue of publication when a third party posts defamatory content on a social media platform. On one view, the matter is posted on Mr Zuckerberg’s platform or some other social media platform. The competing view in a case like *Voller* is that someone posts a defamatory statement on the user’s Facebook page, being impliedly encouraged by that user to do so, or at least in circumstances in which the user is taken to have actively participated in that publication.

Competing perspectives also apply to X/Twitter. For example, Donald Trump’s Twitter page (which one might say was presumptively defamatory) might be perceived to be his domain. However, Twitter had the capacity to modify and remove his and other users’ content, and even to ban Donald Trump from its platform, as it did for a certain period. From that point of view, it is X/Twitter’s domain because it controls it, profits from it, and profits even more if outrageous content attracts more followers and clicks.

These different perspectives are important. They are reflected in stark differences between two circuit courts in the United States about the nature of a social media platform to which I return.

In a different case that concerned whether the First Amendment constrained President Trump’s blocking several users from interacting with his Twitter account, Justice Thomas observed:⁴⁹

“The disparity between Twitter’s control and Mr. Trump’s control is stark, to say the least. Mr. Trump blocked several people from interacting with his messages. Twitter barred Mr. Trump not only from interacting with a few users, but removed him from the entire platform, thus barring *all* Twitter users from interacting with his messages.⁵⁰ Under its terms of service, Twitter can remove any person from the platform – including the President of the United States – ‘at any time for any or no reason’.”

The separate issues of publication and innocent dissemination

The Court of Appeal in *Voller* did not venture into the possible application of the defence of innocent dissemination under s 32 of the *Defamation Act 2005*, and criticised the primary judge for having done so. However, given what the Court concluded about the control that the media organisations had over the contents on their Facebook pages, the media defendants may have struggled to establish that each of them was a “subordinate distributor” for the purpose of s 32(1) of the 2005 Act. Under s 32(2), a person is a “subordinate distributor” of the defamatory matter if the person:

- “(a) was not the first or primary distributor of the matter; and
- (b) was not the author or originator of the matter; and
- (c) did not have any capacity to exercise editorial control over the content of the matter (or over the publication of the matter) before it was first published.”

Without limiting s 32(2)(a), s 32(3) provides examples of when a person is *not* the first or primary distributor of the matter. One is an operator of a communications system by which the matter is made available “by another person over whom the operator has no effective control”.⁵¹

⁴⁹ *Biden v. Knight First Amendment Institute at Columbia University* 141 S.Ct. 1220, 1221 (2021); 593 US (2021); https://www.supremecourt.gov/opinions/20pdf/20-197_5ie6.pdf.

⁵⁰ At the time, Mr Trump’s Twitter account had 89 million followers.

⁵¹ s 32(3)(g).

Section 32(2)(c) raises a difficult issue: when does a party hosting content created by a third party have a *capacity* to exercise editorial control over the content (or over the publication of the matter) before it is first published? For example, if the operator of a website or platform has the *capacity* to block certain content or certain users, to control access to a website, to prioritise where certain content will appear on its platform, to control to whom access is made available (and thereby to whom, if anyone, the material is first published), or to create conditions and rules about the content that is allowed to be posted, then it may have the *capacity* to which s 32(2)(c) refers. If so, it will not be a “subordinate publisher” to whom the s 32 defence may apply.

Problems besetting the concept of publication on social media

Prior to the High Court’s decision in *Voller*, Professor Rolph explained why innocent dissemination is not useful in elucidating the principles of publication. It is a policy-based exculpation of liability for defamation and does not “manifest anything deeply principled about the concept of publication”.⁵² He observed that the case law on whether internet intermediaries are publishers for the purpose of defamation law “is overwhelmingly complex and confusing”.⁵³ He urged a return to basic principle and general notions of tortious responsibility, including the distinction between acts and omissions.

When the defendant engages in a positive act of communication, its liability is strict. When, however, it has failed to deal with a third party’s publication, the defendant is liable only where the defendant has actual notice of the third party’s publication and has the capacity to deal with it, but within a reasonable period of time fails to do so.⁵⁴ In that case, the defendant becomes liable, along with the third party, for the continuing publication.

Professor Rolph concluded by observing that if courts in the 19th Century were able to devise innocent dissemination to respond to the emergence of newspapers and lending libraries, “courts in the 21st century surely have the capacity to develop the common law to address the challenges posed by internet technologies”.⁵⁵

I would like to share that confidence. However, judges are not encouraged to use social media and are not experts in that field. We must wait for cases to come our way, and are limited to facts about new technologies that are presented in the evidence before us. The principle of party presentation that is embedded in the adversarial system means that judges cannot search for evidence and answers in outside material.⁵⁶ We are poorly placed, compared to law reform agencies and expert regulators, to understand the technology or the consequences of new rules that may be developed to address it.

Reforms to Australia’s defamation laws

Prior to the most recent reforms, Australia’s defamation laws had not been reviewed since 2005. This was before the iPhone existed and well before Facebook had achieved mass adoption. Today, a large proportion of defamation cases relate to content published on digital platforms such as Facebook and Twitter. The changing dynamics of media and communication challenge

⁵² D. Rolph, ‘The Concept of Publication in Defamation Law’ (2021) 27 *Torts Law Journal* 1 at 14.

⁵³ At 28.

⁵⁴ At 28.

⁵⁵ At 28.

⁵⁶ PDT Applegarth, ‘Deciding Novel and Routine Cases without Evidence’ (2018) 11(2) *Journal of Tort Law* 173 at 197-200; See generally ‘Judicial Decision-Making and ‘Outside’ Extra-Judicial Knowledge’ (2016); Vol 25 No 3 *Griffith Law Review*.

Australia’s defamation laws. Do current laws properly reconcile freedom of expression and protection of reputation in an increasingly complex, digital world?

In 2018, a review of Australia’s Model Defamation Provisions (“MDPs”) began to determine whether the 2005 laws achieved their intended policy objectives.⁵⁷ It was long overdue. The exponential growth of the internet and the emergence of new technologies meant a significant shift since 2005 in the way that information is published and consumed.

How should defamation law treat social media platforms that are a major source of news and information and a major source of social harm? Are they a kind of “internet intermediary” between content creator and content reader, or a controller of content that has the ability to prioritise selected content, moderate content and remove it?

Stage 1 Review

All Australian jurisdictions, except the Northern Territory and Western Australia, have implemented the Stage 1 amendments agreed to by the Standing Council of Attorneys-General in July 2020.⁵⁸ Those amendments introduced:⁵⁹

- A serious harm threshold to be determined by a judge as soon as practicable before the trial;⁶⁰
- A new public interest defence modelled on section 4 of the UK *Defamation Act* 2013;⁶¹
- A single publication rule which means the limitation period runs from the first publication: this has implications for online content that remains accessible and generates subsequent new publications under the rule in *Dow Jones v Gutnick*.

The serious harm threshold for defamation actions introduced a new element of the cause of action: that the publication of defamatory matter has caused, or is likely to cause, serious harm to the reputation of the person. A party may apply to have the serious harm issue determined by the judicial officer before the trial commences.⁶² This amendment seeks to address concerns about the rising number of defamation actions over “trivial, spurious and vexatious backyard claims”.⁶³ By enabling the serious harm issue to be determined early, the provision aims to filter actions which are likely to result in modest damages awards and reduce the burden on courts dealing with trivial, but time-consuming defamation actions.⁶⁴

⁵⁷ Council of Attorneys-General, *Review of Model Defamation Provisions* (Discussion Paper, February 2019) <https://dcj.nsw.gov.au/documents/about-us/engage-with-us/public-consultations/review-model-defamation-provisions/Final-CAG-Defamation-Discussion-Paper-Feb-2019.pdf>

⁵⁸ LexisNexis, AU Practical Guidance Intellectual Property – Defamation legislative reforms (21 June 2023).

⁵⁹ NSW Government, ‘Review of model defamation provisions’ (16 June 2023) <https://dcj.nsw.gov.au/about-us/engage-with-us/past-consultations/statutory-reviews/review-model-defamation-provisions.html> (“NSW Government Review of model defamation provisions”).

⁶⁰ *Defamation Act* 2005, s10A.

⁶¹ *Defamation Act* 2005, s 29A.

⁶² *Defamation Act* 2005, s 10A(3).

⁶³ New South Wales, *Parliamentary Debates*, Legislative Assembly, 29 July 2020, 2867.

⁶⁴ A. Douvartzidis, C Surman., ‘Is it worth it? A Snapshot of recent defamation cases in Australia’, [2022] (December) *Bulletin of the Law Society of South Australia* 26, 27; *Palmer v McGowan (No 5)* [2022] FCA 893, [523].

One potential application of the serious harm threshold is for an intermediary that is not liable until it knows of the defamatory matter following a complaint, and after that time fails to remove it. It may argue that the real damage was done before it knew about the defamation and therefore any harm from its continued publication was unlikely to be serious.⁶⁵

Stage 2 Review

In December 2022, the Standing Council of Attorneys-General approved in principle the Part A amendments in Stage 2 Review of the MDPs. These amendments seek to address the liability of internet intermediaries for defamatory third-party content.⁶⁶ The amendments are “intended to strike a better balance between protecting reputations and not unreasonably limiting freedom of expression in various circumstances where third parties publish defamatory matter via internet intermediaries.”⁶⁷ Subject to final agreement in 2023, the new provisions are intended to commence from 1 January 2024.⁶⁸

The amendments include:

- Two conditional, statutory exemptions from defamation liability for a narrow group of internet intermediaries, including search engines in relation to organic search results;
- A new innocent dissemination defence for internet intermediaries, subject to a simple complaints process (Recommendation 3B);
- A new court power to make orders against non-party internet intermediaries to prevent access to defamatory matter online;
- A requirement that courts consider balancing factors when making preliminary discovery orders;
- Updates to the mandatory requirements for an offer to make amends for online publications.

In April 2023, the Council approved in principle the Part B Amendments of the Stage 2 Review, which will extend the defence of absolute privilege to police and other complaints-handling bodies.⁶⁹

Conditional exemptions for mere conduits and search engine providers

The Stage 2A amendments propose a conditional exemption for a small group of internet intermediaries including conduits, caching and storage services. They will protect intermediaries that have a passive role in the distribution of content and do not actively engage in publishing, promoting or recommending. The amendments also contain a conditional exemption for search engine providers performing standard search engine functions. The exemption applies where the provider is unable to remove the content and does not have an interest in publication or a relationship with the originator. The exemption does not extend to sponsored content or snippets.

⁶⁵ *Tamiz v Google* [2013] EWCA Civ 68 at [50].

⁶⁶ [NSW Government Review of model defamation provisions.](#)

⁶⁷ Standing Council of Attorneys-General, ‘Communique’, (9 December 2022)

<https://www.ag.gov.au/about-us/publications/standing-council-attorneys-general-communications>.

⁶⁸ [NSW Government Review of model defamation provisions.](#)

⁶⁹ Standing Council of Attorneys-General, ‘Communique’, (28 April 2023) <https://www.ag.gov.au/about-us/publications/standing-council-attorneys-general-communications>.

Innocent dissemination defence subject to complaints notice process

The amendments will introduce a new innocent dissemination defence for internet intermediaries. It involves a complaints process. It aims to ensure that internet intermediaries are not liable for third party content where they are merely a subordinate distributor and lack knowledge of the content.⁷⁰ Once an internet intermediary receives a complaint it must take reasonable steps to remove or otherwise prevent access to the matter within a 14 day period.⁷¹ The defence will apply to a “digital intermediary” which is “a person, other than an author, originator or poster of the matter, who provides an online service in connection with the publication of the matter”.⁷² This is anticipated to include anyone who uses a social media platform to create and administer a public page.⁷³ Therefore, the effect of *Voller* may be reversed by the amendments if such a media defendant removes the defamatory content within 14 days of complaint.

An exemption to s 235(1) Online Safety Act

Schedule 5, clause 91 of the *Broadcasting Services Act 1992* (Cth) (now repealed) limited liability for internet content hosts (“ICHs”) and internet service providers (“ISPs”) from State and Territory laws for third party content where they are not aware of the nature of the content. This limit was introduced in 1999. It required the ICH or ISP to have knowledge of the defamatory matter before they could be held liable.⁷⁴ The legislature intended the provision to apply to a narrow group of intermediaries. ICHs and ISPs are treated as distinct from content providers and users who have actual control over the publication and accessibility of content.⁷⁵

The *Online Safety Act 2021* (Cth) was introduced to regulate ISPs and the dissemination of harmful content.⁷⁶ Section 235(1) provides a similar protection for Australian hosting service providers (HSPs) and ISPs as was previously provided by clause 91 of the *Broadcasting Services Act*. It essentially provides that a State or Territory law or rule of common law or equity has no effect to the extent that it:

- subjects HSPs and ISPs to liability for content where it is not aware of the nature of such content; or
- requires HSPs and ISPs to monitor, make inquiries about or keep records of online content that it hosts.

⁷⁰ Meeting of Attorneys-General: Stage 2 Review of the Model Defamation Provisions – Part A: liability of internet intermediaries for third-party content – Summary Paper: Model Defamation Amendment Provisions 2022 (Consultation Draft), <https://dcj.nsw.gov.au/documents/about-us/engage-with-us/public-consultations/review-model-defamation-provisions/summary-paper-for-draft-part-a-model-defamation-amendment-provisions-2022.pdf>.

⁷¹ Meeting of Attorneys-General: Stage 2 Review of the Model Defamation Provisions – Part A: liability of internet intermediaries for third-party content – Background Paper: Model Defamation Amendment Provisions 2022 (Consultation Draft), at 42 <https://dcj.nsw.gov.au/documents/about-us/engage-with-us/public-consultations/review-model-defamation-provisions/background-paper--for-draft-part-a-model-defamation-amendment-provisions-2022.pdf> (“Stage 2 Review Background Paper”).

⁷² Stage 2 Review Background Paper at 43; definition of “digital intermediary” in MDA Amendment Provisions, 2022.

⁷³ A. Meade, ‘Defamation reforms: Australian media may not be liable for Facebook comments in future’, *The Guardian*, (14 December 2022) <https://www.theguardian.com/media/2022/dec/14/defamation-reforms-australian-media-may-not-be-liable-for-facebook-comments-in-future>.

⁷⁴ A. Latu, ‘Online intermediaries and defamation “safe harbours” in Australia and New Zealand: Content hosts, Facebook comments and contretemps’ (2021) 28(2) *Tort Law Review* 61 at 68.

⁷⁵ At 72.

⁷⁶ P. George, *Defamation Law in Australia* (LexisNexis, 4th ed, 2022) at 608.

During the Stage 2 review, stakeholders raised concerns about the lack of clarity around the operation of section 235(1) and the liability of internet intermediaries in defamation. The review highlighted the need to clarify which intermediaries are protected and what constitutes ‘awareness’ of the content.⁷⁷ Concern was expressed that the proposed amendments to defamation laws would encounter the s 235(1) prohibition on requiring providers to monitor, make inquiries and keep records.⁷⁸ In December 2022, the Standing Council of Attorneys-General agreed to consider the desirability of an exemption from s235(1) for defamation law.

Other legislative developments

Reforms to Australia’s defamation law to address challenges posed by social media have been a long time in the making.⁷⁹ In the meantime, developments have occurred in judge-made law, including two significant decisions by the High Court in *Voller* (2021) and in *Google LLC v Defteros* (2022), and the *Online Safety Act 2021* (Cth) has been enacted. Laws have been enacted to address specific evils, like the kind of material that was posted by the individual responsible for the Christchurch massacre,⁸⁰ and the non-consensual sharing of intimate images over a carriage service.⁸¹

Other jurisdictions have developed new laws addressing the dissemination of harmful content. They include broad legislative schemes, like the *Online Safety Act*, to deal with the dissemination of a variety of harmful content. Currently, the UK is considering an *Online Safety Bill* to address “harmful conduct”. It aims to regulate companies rather than the individuals who post material. Again, it addresses a wide range of harmful conduct, not just hate speech. It includes lawful communications that may be harmful to children. If passed, it will attempt to regulate internet entities. Leaving aside issues of principle about subjecting such entities to state control, informed commentators wonder whether any regulator will be able to do what it is expected to do.

Overseas responses to misinformation and disinformation

As a community we are confronted by the broad issue of misinformation and disinformation. False defamatory matter is one legal subcategory of misinformation and disinformation. Legal scholars have addressed how we should deal with fake news. Professor Cass Sunstein’s work “Liars”⁸² is a good example.

Early in 2018, the European Commission created a group of experts to advise on how to counter fake news and online disinformation. Its report recognised that any measures should not violate fundamental principles of freedom of expression or undermine the technical functioning of the internet. The expert group supported targeted, non-regulatory approaches to the problem, rather than broad regulatory interventions. One recommendation was that steps be taken to ensure digital media companies are more transparent about news production and distribution. One aim was to support users to identify who is behind a certain type of information, their motivations and their sources of funding. The use of “source transparency indicators” is designed to help users better discern between credible journalistic content and disinformation.

⁷⁷ [Stage 2 Review Background Paper](#).

⁷⁸ [Stage 2 Review Background Paper](#) at 44-46.

⁷⁹ For example, in 2013 the UK created a new defence for operators of websites who did not post the statement complained of: *Defamation Act 2013* (UK) s 5.

⁸⁰ In 2019, Div 474 was inserted into the *Commonwealth Criminal Code* to prohibit the sharing of abhorrent, violent material and to regulate social media platforms.

⁸¹ In 2018, the Commonwealth introduced such a law and State laws also prohibit such conduct, known by the term ‘revenge porn’. See P. George ‘*Social Media and the Law*’ at [8.92]-[8.103].

⁸² C. Sunstein, *Liars: Falsehoods and Free Speech in an Age of Deception* (Oxford University Press, 2021).

Other steps include educating the public as users of platforms and improving media and information literacy.

The federal government's plans to address online misinformation

In late June 2023, the federal government announced proposed legislation to boost the powers of the Australian Communications and Media Authority (ACMA) to compel tech companies to keep records about misinformation and disinformation, and to provide them to ACMA, when requested. The aim is to enable ACMA to monitor efforts and require digital platforms to do more. Under the proposed powers, ACMA might request an entity to deliver a Code of Practice about measures to combat misinformation and disinformation on digital platforms, which ACMA could register and enforce. The proposed powers also would allow ACMA to create and enforce an Industry Standard in the event a Code of Practice was deemed ineffective in combatting misinformation and disinformation on digital platforms. An Industry Standard is a stronger form of regulation than a Code of Practice.

Importantly, ACMA will not enjoy the power to request specific content or a post be removed from digital platform services. These proposals follow ACMA's June 2021 report on the adequacy of digital platforms' disinformation and use quality measures.⁸³

The approach of equipping a regulator to compel the provision of information from the operator of a digital platform follows the approach adopted by the *Online Safety Act* in 2021. The 2021 law gives the eSafety Commissioner the power to compel companies to explain what they are doing to control certain forms of harmful content, such as child sexual abuse material and hate speech. The 2021 Act gave the eSafety Commissioner power to compel internet service providers to block "abhorrent, violent material" in certain circumstances. Unlike that law, the proposed laws about misinformation and disinformation will not equip the regulator to direct that material be blocked. This reflects the importance of freedom of speech interests. It also may reflect the practical reality that Australian regulators lack the ability to control social media sites that are based overseas.

An international perspective

Difficult policy decisions are made by sovereign nations about the types of regulations that should be imposed on major and minor players on social media. Countries like Australia may take their lead from proposals that emerge from Europe. However, each country has to enact laws that are correct in principle and likely to work in practice to achieve their aims of controlling misinformation and disinformation, including indefensible defamations, being communicated by social media platforms.

An interesting example of one nation going alone concerns an unrelated area of social media platforms' use of content created by others, including legacy media. We had the confrontation between the Australian Government and tech giants in 2021 that ultimately resulted in deals being struck. That deal was struck only after Facebook temporarily blocked Australian users from sharing news on its site. In theory, if a tech giant finds the laws or regulations passed by a country too onerous, it can threaten to block a country's users from its platforms.

⁸³ Australian Communications and Media Authority, *A report to government on the adequacy of digital platforms' disinformation and news quality measures* (Report, June 2021) <https://www.acma.gov.au/sites/default/files/2021-11/Adequacy%20of%20digital%20platforms%20disinformation%20and%20news%20quality%20measures.pdf>.

In 2019, one of the founders of the internet, Sir Tim Berners-Lee, urged governments to translate laws and regulations for the digital age so as to ensure that “markets remain competitive, innovative, and open”. He argued that governments have a responsibility to protect people’s rights and freedoms online.⁸⁴ He reflected on occasions when governments around the world have worked together, passing the *Universal Declaration of Human Rights* and laws to govern the sea and outer space.

The ideal would be to have consistent laws around the globe to regulate a global industry. Nations may be able to reach a consensus about marine pollution, the international carriage of goods and the enforcement of arbitration awards. However, they seem unlikely to negotiate a Convention on regulating social media in the near future. There are fundamental differences between nations about reconciling freedom of speech and other values, like privacy, and the extent of permissible State regulation of content.

American lawmakers seem incapable of deciding what should be done about s 230 of their 1996 Act, let alone fall into line with some international rule of law about social media.

Unless and until governments agree on laws and regulations affecting social media, the laws of significant polities will influence conduct. This may be seen with the development of privacy laws in Europe. In practice, tech giants are forced to comply with them if they wish to operate in Europe. In that way, European standards flow through to the rest of the world.

The architecture of local laws about internet intermediaries’ liability for third-party content

Returning to local problems and local solutions, the traditional analysis, based on old media involving journalists, editors, publishers, printers, distributors and other communicators of defamatory content, was to ask:

- (1) Is someone a publisher and, therefore, *prima facie* liable for defamation?
- (2) If so, are they the kind of subsidiary publisher to whom an innocent dissemination defence, common law or statutory, might apply?

In 2005, Australia developed a uniform, innocent dissemination defence. However, its application to internet intermediaries is uncertain. If a social media entity is a “subordinate publisher” within the meaning of s 32(2) *Defamation Act* 2005, as clarified by s 32(3), the defence requires it to prove more than lack of knowledge. It must prove that it ought not reasonably have known the matter was defamatory *and* that its lack of knowledge was not due to any negligence on its part. This last element raises the issues of the adequacy of the defendant’s monitoring and moderation practices and the incompatibility of such a State law with federal law that neutralises any State law’s requirement for an HSP or ISP to monitor content that it hosts.

The policy issue is whether “internet intermediaries” (however defined) should enjoy some additional defence or statutory immunity that is tailored to suit their activities.

⁸⁴ T. Berners-Lee, ‘30 years on, what’s next #ForTheWeb?’ (World Wide Web Foundation, 12 March 2019) <https://webfoundation.org/2019/03/web-birthday-30/>.

If we had our time over, rather than tack on additional defences for internet intermediaries, we might approach the general issue as follows:

- (1) Is someone a publisher and, therefore, *prima facie* liable for defamation?
- (2) Regardless of that issue, should they have some kind of statutory immunity from liability for defamation or, at least, an innocent dissemination defence?
- (3) If so, what should be the conditions upon which such an immunity or defence depends?
- (4) Are “internet intermediaries” different to other disseminators?
- (5) Is an entity like Facebook that can moderate content on its platform a different kind of “internet intermediary” to one that passively communicates content?

Difficult architectural decisions must be made. Should immunities or defences be defined by:

- what the party is (e.g. “an internet content host”);
- what the party does; or
- a definition that excludes entities that might be variously described as a principal or primary publisher, originator or controller of editorial content from enjoying the immunity or defence.

How are internet intermediaries different?

A threshold question is whether certain internet intermediaries are publishers. As the legacy media defendants that lost the argument in *Voller* know to their expense, not every party that hosts third-party content can avoid being treated as a publisher. Those entities that seemingly are not publishers include the kind of ISPs that supply internet carriage services to the public, caching services and services that enable the storage of data. That narrow class of intermediary that does not select the intended recipient or “encourage, edit or promote the matter” will enjoy statutory exemption from defamation liability under Recommendation 1 of the Attorneys-General proposed amendments.

Current laws, particularly the law that was passed in 1999 and remains on the books, place internet intermediaries in a different category to traditional subordinate publishers that enjoy the innocent dissemination defence. Under the 1999 law some, but not all, internet content hosts are relieved of any obligation to monitor, including that impliedly imposed by the conditions of innocent dissemination defences.

Internet intermediaries, however defined, present different challenges and opportunities than those presented by most old-fashioned subordinate publishers or facilitators, such as the distributors of newspapers, as well as some familiar problems. Once a newspaper boy selling newspapers at the traffic lights had sold the paper to a motorist, the damage was done. The same applied to printers and most other distributors. They had limited power to retrieve the newspapers or magazines they had sold. Libraries and bookstores may be able to take books off their shelves. But once the book has been lent or sold, the damage has been done. The innocent dissemination defence insulated the subordinate publisher from liability for past damage, but encouraged it to not cause further damage once it was aware of the defamatory matter.

The rise of social media and easy access by citizens around the world to material hosted on the internet means that take-down remedies assume importance for those who are defamed.

The Stage 2A defamation law reforms present the issue of liability for hosting and communicating third-party content as one that ought to be resolved as far as possible between the defamed and the originator of the content. However, the natural concern and grievance of those who are indefensibly defamed do not lie simply with the originator. The grievance often is against the entity that facilitates the defamatory content being communicated to the world. This was the case with plaintiffs who sued newspapers over letters to the editor.

Current law renders a primary publisher, like a newspaper, liable for the defamatory statements of a third party whose publication it facilitates on its pages. The same is true of broadcasters who communicate defamatory statements by third parties over the airwaves. Talk-back radio stations have a seven-second delay button to censor defamations before they are broadcast. The difference between those traditional publishers of third-party content and most internet content hosts that communicate defamatory matter on their platforms and services is that most internet content hosts do not monitor content that is posted at the time it is posted, let alone before it is published.

Describing newspapers and broadcasters as primary publishers because they exert editorial control over third-party content before the point of mass publication and social media platforms as subsidiary publishers or intermediaries because they do not, may seem harsh to many. It privileges a new form of media that has gained its ascendancy and power as a major source of news and information because its business fails to employ individuals or even machines to screen harmful defamatory content in advance or even to monitor much of the content it hosts shortly after posting and before a complaint arrives. Privileging social media is, however, the prevailing public policy.

Rather than heroically argue that the social media platforms against which they compete should be subject to the same liabilities to which they are subject, the media defendants in *Voller* sought and will be granted the privileges associated with being an internet intermediary when a third party posts defamatory content to their site and they host it.

There is an obvious difference between the old-fashioned newspaper that decides to publish a letter to the editor in its print edition and, if it has an online edition, to publish the same content in its online edition. An editor makes a decision to publish a third party's content, knowing of its content. Third-party content on social media is different. Third parties, literally from anywhere in the world, can post content on a media organisation's social media page if it is a public page. They also might post items where readers' comments are open beneath particular articles. One might hope that there is the functional equivalent of a "Letters to the Editor" sub-editor monitoring these posts. However, the billionaire owners of legacy media have sacked the staff who might monitor that kind of content. As a result, legacy media lack the resources and the people to moderate the defamatory content that is posted on their websites and online platforms. Instead, they rely upon people who are harmed by such content to complain about it and thereby bring it to their attention.

Legacy media that host defamatory content on their social media pages are to be placed on the same legal footing as other so-called "internet intermediaries" under new Australian laws. They will be relieved of any implied obligation to monitor content under current innocent dissemination defences that are lost if a publisher's ignorance of the defamatory content is due to a lack of care.

The new defence will be available to a “digital intermediary”. That includes a person, other than an author, originator or poster of the matter, who provides an online service in connection with the publication of the matter. The broad definition of “online service” encompasses all of the internet intermediary functions within the scope of the review, and so includes social media platforms, review websites and forum administrators.

Rather than adopt the Recommendation 3A proposal for a safe harbour defence for digital intermediaries, subject to a simple complaints notice process, the Council of Attorneys-General adopted Recommendation 3B. The idea is to introduce an actual notice requirement in an innocent dissemination defence for internet intermediaries, subject to a simple complaints notice process. This draws on a similar model to that in the UK.

Section 5 of the 2013 UK Act created a new defence for “operators of websites” who do not “post” the statement on its website. Save where the operator acts with malice, the UK defence is unconditional in respect of identifiable authors. In the case of unidentified authors, the defence is conditional on compliance with regulations about actions required of an operator, including action about the identity or contact details of the person who posted the statement and action relating to its removal. Under s 5(12), the fact that the operator “moderates” the statements of others posted on the website does not of itself defeat this defence.

Defamation legislation of the kind that has existed in the UK for a decade, and which is finally pending in Australia for internet content hosts, is necessary because internet content hosts are different to traditional media that host third-party content after making an editorial decision to do so, even a rapid decision by a talk-back radio producer who decides not to push the seven-second delay button.

The limits on legislation

Policy decisions need to be made as to whether liability should reside with the originator of the content, the host and mass-communicator of that content, or both. If certain hosts of harmful content are to have a conditional immunity from liability or access to a tailor-made “innocent dissemination” defence, who should those hosts be and what conditions should be set?

A perennial problem is to define terms like “internet content host” (cl 91 BSA), “Australian hosting service provider” (OSA), “operator of a website” (2013 UK Act), or “digital intermediary” (MDA Amendment Provisions 2022, s 4). Any definition that excludes parties that host third-party content over which they exercise editorial control may be interpreted to exclude a party that moderates content, including by automatically removing or blocking certain words or monitoring for offensive or defamatory statements. A broad definition such as “a person, other than an author, originator or poster of the matter, who provides an online service in connection with the publication of the matter” (the meaning of “digital intermediary” in the proposed *Defamation Act* amendments) captures a wide class, including content hosts that exercise editorial control over content.

As occurred in 1999 in defining an “internet content host” and in 2005 in defining a “subordinate publisher”, legislators adopt definitions that are intended to fit the actors and activities that then exist and that are foreseeable. Legislation has to anticipate technological developments and not be unduly specific. On the other hand, a reasonable degree of legal certainty is required about the entities or activities that fall within a statutory definition. Unanticipated technological developments and new business models arrive years later.

One example is the advent of social media influencers. They spread misinformation about things like the health benefits of a product they promote. They may promote a certain product by saying nasty, misleading or defamatory things about a competitor's product. The social media influencer is in a business relationship with the manufacturer of the product they promote. How do laws, including prohibitions against misleading and deceptive conduct and the law of defamation, address the liability of such an undisclosed principal?

Decisions about the architecture of new laws to regulate social media, including the precision or breadth of definitions, arises in a particular legal context. A certain scheme may be appropriate in the context of criminal laws prohibiting the dissemination of child pornography. A lighter regulatory approach may be called for where the content concerns communication about government and political issues that often includes defamatory matter. General online harm laws may cover a multiplicity of harms. A regulatory scheme may be appropriate to communications of child pornography, but entirely inappropriate to communications about matters of legitimate public interest, particularly where the content is generated by professional journalists.

This leads some legislators to create exemptions or carve-outs for journalistic content, so as to avoid the heavy hand of State censorship by a government agency ordering professional journalists and their editors to take down material. However, a journalistic exemption, however described, creates its own drafting problem of defining what journalism is and who practises it. Many so-called "citizen journalists" claim to be practising journalism. So do many lonely bloggers. One loosely worded definition of "journalistic enterprise" in a Florida statute includes an entity that posts more than 100 hours of video and has more than 100 viewers per year. According to the 11th Circuit Court of Appeal, this qualified PornHub to be a "journalist enterprise".

Encouraging moderation

In developing new laws, regulations and codes of practice for internet intermediaries, it is worth reflecting on how defamation law historically has approached the issue of liability for third-party content. Central to the law of defamation is the republication rule that renders a publisher liable for republishing the defamatory words of another. Centuries ago, the law provided that communicators should be *prima facie* liable for publishing the words of others, because those communicators were responsible, along with the originator, for the harm caused to the indefensibly defamed.

The common law has adhered to traditional definitions of who is a publisher. Liability is strict and does not depend on knowledge or a failure to take reasonable care. A law of strict liability was untenable and unfair on certain innocent disseminators who exercised care and were not aware of the defamatory content. Hence, the defence of innocent dissemination.

The rise of the internet brought new challenges. Those of us who were alive at the time remember how online chat rooms developed for persons in certain fields of science and by small community groups and clubs. A chat room might be created by some well-intentioned group and someone was selected to act as its moderator. More than 20 years ago, cases like *Godfrey v Demon Internet* addressed these issues. They suggested that hosts of defamatory content in discussion groups or chat rooms were liable for defamation, at least after they knew of the defamatory content, typically when a complaint was made about it. Even a few decades later, the idea that the moderator of a community-based social media page should be liable for the defamatory content that is posted on it by a stranger, is confronting to some, surprising to others, and costly for moderators.

Having regard to the internet's modest beginnings and its potential for social good, laws understandably sought to encourage its development and to provide protection for internet

service providers and certain internet content hosts. IT contractors who set up and rather passively hosted the website of a business, a group of professionals, or an individual hosting the contents of a discussion group on their server were thought deserving of protection as an “internet content host”.

Most citizens would think it a good thing that persons who host content and have the ability to remove harmful content might monitor content and not increase their risk of legal liability in doing so. The policy is that internet content hosts should not be discouraged from monitoring the content that they host by the risk that by exercising the power to monitor, moderate or remove content, they will be transformed from being a passive, non-publisher into a publisher with *prima facie* liability, that has to prove an innocent dissemination defence.

The regulatory response in different countries was to encourage moderation. Internet content hosts should enjoy additional protections beyond that provided to traditional innocent disseminators if, for some reason they failed to identify defamatory or other harmful content and, therefore, did not know about it.

In Australia in 1999 they were given a statutory safe harbour until they knew about the content. Federal law neutralised any State or Territory law that obliged them to monitor content. Australian law created a safe harbour for internet content hosts to avoid liability for content of which they were not aware. US law went to an extreme. Social media platforms have no obligation imposed upon them to remove harmful content even after they are alerted to it.

US law has conferred a monumental, competitive advantage on giant social media platforms like Facebook and X/Twitter by insulating them from legal liability. They are the beneficiaries of a law that its supporters claim gave us the internet. It was, however, a law principally directed at online pornography.

American-based social media giants are largely beyond the reach of Australian law, insofar as they do not have a legal presence in the jurisdiction. Litigants try and fail to sue their Australian-based subsidiaries which neither control nor host the content in Australia. Claimants wishing to sue overseas-based companies need court leave to serve outside Australia under “long arm” jurisdiction rules.

In the possibly unlikely event that an American-based communication giant, whether it be a social media platform or the old-fashioned *New York Times*, has a defamation judgment entered against it in Australia, American laws effectively prevent such a judgment from being enforced in the US. The *Speech Act 2010* (US) protects them against foreign defamation judgments in places when the law provides less legal protection for free speech than under US law.

Potential civil law liability as an incentive to moderate

The regulation of social media in the US leaves moderation and removal of content to the entities that enjoy the broad immunity given to them by s 230.

In Australia, legal uncertainty exists over which internet content hosts attract the statutory protection of the *Online Safety Act*, s 235. However, a “hosting service provider” is encouraged by the threat of liability to moderate content at least after it becomes aware of it. If an entity does not enjoy the protection from civil liability conferred by s 235, then its only real form of protection is the innocent dissemination defence under s 32 *Defamation Act 2005*. That defence provides an incentive to create monitoring systems to ensure that a lack of awareness of defamatory content is not due to the entity’s negligence.

An enduring problem - to censor or not to censor?

Assuming always that an entity is taken to be a publisher, laws that expose such an entity to liability for defamation after it becomes aware of the defamatory content, present a vexing public policy issue. To minimise legal risk, such an entity may remove the content as soon as it becomes aware of it, typically after receiving a complaint. This carries a risk of excessive self-censorship. The risk of legal liability may prompt the removal of content that is true and that should be published in the public interest.

The existing defence of innocent dissemination softens a tort of strict liability into something akin to a negligence-based form of liability for dissemination. The defence depends upon the defendant proving that its lack of knowledge that the matter was defamatory was “not due to any negligence on the part of the defendant”.⁸⁵ The defendant ceases to have legal protection from defamation once it is on notice. It then must decide whether to further publish the defamatory matter.

The proposed Stage 2 Part A laws create what is described as a simple complaints process. Once the internet intermediary has received a complaint notice, it must prevent access to the matter within 14 days in order to be able to rely upon the new statutory defence.

In the past, some highly-motivated individuals, upset by what a journalist or author wrote about them, would focus their legal resources on a distributor, such as a bookstore or newsagency chain. By threatening legal action against the distributor they hoped to induce the distributor or subordinate publisher to take the book or magazine out of circulation.⁸⁶

The policy issue of what liability, if any, should be imposed on entities that host the defamatory content of third parties is not new. It is a problem that has confronted newspapers over letters to the editor. Responsible newspaper editors, conscious of the need to publish different, dissenting and diverse opinions, publish letters to the editor. They provide citizens with the opportunity to publish harsh and hurtful criticism, including statements that the editor may believe to be untrue.

In 1960, the *New York Times* accepted a paid advertisement entitled “Heed Their Rising Voices” from civil rights activists. It was not a letter to the editor. It was a paid advertisement. The *New York Times* was not a secondary or subordinate publisher. In contemporary parlance, it was a “content host” in publishing the advertisement. Its advertising manager made a decision to publish the advertisement. He might have chosen not to do so because it was a controversial item that criticised the conduct of police in Alabama. It would have been safer to publish an advertisement for Kellogg’s Cornflakes. The *New York Times* was punished by a southern jury which awarded the unnamed police commissioner of Montgomery, Alabama, \$500,000. The case led to the creation of a constitutional defence in *New York Times Co. v Sullivan*.⁸⁷

It is easy to say that the *New York Times* was on the side of the angels in publishing what it did at that time. It may have faced a harder choice to accept a paid advertisement from the Ku Klux Klan.

Media today face similar choices about hosting the defamatory content of third parties. Should its decision to publish content that attacks the vulnerable or the disadvantaged, like victims of racism or sexual assault, be left to its editorial decisions and good judgment? Are media entities encouraged to make good editorial decisions by the threat of civil liability?

⁸⁵ *Defamation Act 2005*, s 32(1)(c).

⁸⁶ See *Goldsmith v Sperrings Ltd and various distributors* [1977] 1 WLR 478.

⁸⁷ 376 US 254 (1964).

Can it be said that the *Australian Financial Review*, in recently publishing a disgraceful cartoon in the form of a paid advertisement, was some kind of neutral content host? If not, is the position any different for an online social media platform that does the same thing?

Ultimately, the issue is how publishers, big and small, online and legacy, make decisions about publishing harmful and offensive material, including defamatory matter. One aspect is how the law, both the general law and statute law, should influence those decisions.

A need exists to protect the vulnerable and the defamed by not publishing indefensible third-party content, or to remove it once that content comes to the publisher's attention. On the other hand, there is a need to avoid excessive self-censorship prompted by the risk of legal liability.

Our views about how the law should reconcile these competing public interests depends somewhat on our perceptions of content hosts and third-party content creators. We might imagine the content host as being a profit-driven, nasty media empire that encourages hatred and division, sacks journalists and relies on free content created by third parties. Instead, we can imagine the host as being in the heroic tradition of the *New York Times* or one of the high quality Fairfax publications of the 1980s that practised investigative journalism.

As for the originators of the third-party content, we might imagine the propagator of a racist, conspiracy theorist who believes the 2020 US Presidential election was stolen, or some other mischievous untruth. We can imagine some malicious keyboard warrior who has no fear about being sued because he has no money. By contrast, it is possible to imagine the originator being a well-intentioned, anonymous whistle-blower, intent on exposing official corruption, or a foreign student who wishes to expose the actions of a brutal, foreign government.

The law has to address all of these stereotypes about third-party content creators and those who host and communicate their content.

There is a danger in laws encouraging some publishers to give up the identity of their sources too readily as the price to be paid to avoid liability. Other publishers simply will not do so.

The proposed Stage 2 Part A reforms attempt to ensure that the fight is between the defamed complainant and the originator, so far as possible. Yet, the originator may be a person who has no assets and it will be a small comfort to the harmed to know who they are. The injured, defamed individual is inclined to say, as Mr Voller did, that the real harm was not done by the originator who created the nasty content, but by the media entity that hosted it and communicated it to the world.

The Stage 2 Part A Reforms are less than perfect. However, no one can expect perfection in an area of such complexity. An aggrieved, defamed individual may think that allowing a wealthy media entity 14 days to decide whether to take down the offending material is far too long. Enormous additional damage can be done in 14 days. A content host, large or small, will say that it requires that time to investigate the complaint and make appropriate decisions about whether to remove the content, lest there be hasty and excessive censorship.

Take down orders

An important remedy for the indefensibly defamed is to have material taken down from public access as soon as possible. The reforms propose an interesting solution. They provide for defamation laws to be amended to provide that where a court grants an interim or final order or judgment for the complainant in an action for defamation, the court may order a person who is not a party to remove, block or disable access to the online matter within the scope of such order or judgment. Naturally, notice must be given to the non-party before the order is made.

Again, individuals who are harmed by damaging, defamatory material may have to wait some time for such a remedy to be granted against a content host. They first have to sue the originator and obtain an order. Still, this new form of remedy will be welcomed by those who are able to get into court quickly and obtain an interim injunction against a malicious keyboard warrior.

Defamatory meaning

In any defamation case there are three key questions, before one begins to consider innocent dissemination and other defences:

- (1) Is the defendant a publisher?
- (2) What did it publish and about whom did it publish it?
- (3) Would the publication convey the defamatory meaning alleged by the claimant to an ordinary, reasonable reader?

The third issue concerns an interpretation by a hypothetical reader, about whose habits and thinking has been written in the context of traditional media.⁸⁸ We now have to consider the habits and thinking of an ordinary, reasonable, social media user.

Many decisions have established that the more casual, speech-like characteristics of exchanges on social media sites are relevant to the determination of meaning.⁸⁹ In *Stocker v Stocker*,⁹⁰ Lord Kerr stated that it was:

“the court’s duty to step aside from a lawyerly analysis and to inhabit the world of the typical reader of a Facebook post. To fulfil that obligation, the court should be particularly conscious of the context in which the statement was made...”

He continued:⁹¹

“The fact that this was a Facebook post is critical. The advent of the 21st century has brought with it a new class of reader: the social media user. The judge tasked with deciding how a Facebook post or a tweet on Twitter would be interpreted by a social media user must keep in mind the way in which such postings and tweets are made and read.”

The court’s task is to ascertain how a typical reader would interpret a message and:⁹²

“That search should reflect the circumstance that this is a casual medium; it is in the nature of conversation rather than carefully chosen expression; and that is pre-eminently one in which the reader reads and passes on.”

⁸⁸ *Reader’s Digest Services Pty Ltd v Lamb* (1982) 150 CLR 500, 506; *Farquhar v Bottom* [1980] 2 NSWLR 380 at 386; *Amalgamated Television Services Pty Ltd v Marsden* (1998) 43 NSWLR 158 at 165-166; *Radio 2UE Sydney Pty Ltd v Chesterton* (2009) 238 CLR 460 at 477-478.

⁸⁹ R. Parkes et al (eds), *Gatley on Libel and Slander* (Sweet & Maxwell, 13th ed, 2022) at 116 [3-030].
⁹⁰ [2019] UKSC 17; [2020] AC 593 at [38].

⁹¹ At [41].

⁹² At [43].

In *Monroe v Hopkins*,⁹³ Warby J noted that Twitter is a “conversational medium” and that an:⁹⁴

“impressionistic approach must take account of the whole tweet and the context in which the ordinary reasonable reader would read that tweet. That context includes (a) matters of ordinary general knowledge; and (b) matters that were put before that reader via Twitter.”

In the “Wagatha Christie” case involving footballers’ wives and girlfriends, *Vardy v Rooney*,⁹⁵ Instagram posts were in issue. Warby J observed of this medium:⁹⁶

“People will tend to scroll through messages relatively quickly. The reader’s reaction to a post is impressionistic and fleeting. The reader is likely to absorb the essential message quickly, before moving on. Readers of social media do not have advocates beside or in front of them, making arguments about what a tweet means.”

A claimant must sue over the whole of a publication, not parts taken in isolation, and the judge or jury must have regard to the context in which words appear in deciding the meaning they would convey to an ordinary, reasonable reader. This does not prevent a claimant suing over a snippet or a tweet and pleading it as the publication complained of. This is illustrated by *Hockey v Fairfax Media Publications Pty Ltd*,⁹⁷ in which the newspaper article did not convey the defamatory meanings alleged. However, the claimant was able to successfully sue over tweets and a poster that stated “Treasurer for Sale” or “Treasurer Hockey for Sale”. He failed in his claim over the article. Therefore, it was a minor victory, with Mr Hockey being awarded only 15 per cent of his costs. The fact that the tweets gave a hyperlink to an article that altered and qualified the tweet’s meaning, being an article that was not defamatory, did not mean that the tweet was not itself defamatory. Many readers of the tweet would go no further.

The law of defamation developed on the assumption that the ordinary, reasonable reader read the whole of a newspaper article, often placing greater emphasis on headlines and opening paragraphs.⁹⁸ In addition to assuming that readers read the same thing, defamation law adopts the “single meaning” rule, despite the reality that reasonable people may derive different and inconsistent meanings from the same words. The “single meaning” rule has been called into question,⁹⁹ and in *Falter v Alzmon*,¹⁰⁰ Nicklin J questioned its application to the online world.¹⁰¹

“*Charleston v News Group Newspapers* comes from a different era where print copies of newspapers were essentially the main medium through which people were defamed. In such cases it was relatively straightforward, given that the totality of what was provided to the reader was readily available, to treat the ordinary reasonable reader as having read the entirety of an article including its text, headline, text, furniture, things like that. The Internet provides a degree of challenge to that orthodoxy because it is possible to set out in on-line publications many hyperlinks to external material. It is perhaps unrealistic to proceed on the basis that every reader will follow all the hyperlinks, but everything depends upon its context. For example, if in a single tweet there is a single statement

⁹³ [2017] EWHC 433 (QB); [2017] 4 WLR 68.

⁹⁴ At [35].

⁹⁵ [2020] EWHC 3156 (QB).

⁹⁶ At [18].

⁹⁷ [2015] FCA 652.

⁹⁸ *Gatley* at [3.031]-[3.032].

⁹⁹ D. Rolph, *Defamation Law* (Lawbook Co., 2016) at [6.140].

¹⁰⁰ [2018] EWHC 1728 (QB).

¹⁰¹ [2018] EWHC 1728 (QB) at [11]-[13]. See also *Greenstein v Campaign Against Antisemitism* [2019] EWHC 281 (QB) at [18].

that says, ‘X is a liar’ and then a hyperlink is given, it is almost an irresistible inference to conclude that the ordinary reasonable reader would have to follow the hyperlink in order to make sense of what was being said. At the other end of the spectrum, a very long article could contain a very large number of hyperlinks. Only the most tenacious or diligent reader could be expected to follow every single one of those hyperlinks. Such a reader could hardly be described as the ordinary reasonable reader. How many links any individual reader would follow would depend on an individual’s interest in or knowledge of the subject matter or perhaps other particular reasons for investigating each of the hyperlinks in question.

It therefore does not seem to me to be possible to put forward a hard and fast rule that hyperlinks imbedded in an article that is complained of should be treated as having been read by the ordinary reasonable reader.”

*Gatley on Libel and Slander*¹⁰² reviews certain cases. In some, hyperlinked material was found to form part of the context of the words complained of, while others ruled that the statements sued on should be treated in isolation, with links being further, not required reading.

In short, the law governing interpretation of online publications is complicated.

Search results and snippets

An intriguing issue is the meaning that might be derived from Google search results that are published in response to an entry of search terms into Google’s search engine. For example, what defamatory meaning, if any, is reasonably conveyed by typing the words “Judge” and “paedophile” into the search engine, whereupon it produces an image and the name of a judge. Is such a search result capable of imputing that the judge is a paedophile or suspected of being a paedophile, or would any reasonable reader of that result assume that the judge’s name and image appeared because he was the trial judge in a proceeding against a paedophile? Long-standing principles of defamation law have had to be adapted to a different world.

The display of a “snippet” from a Google search engine was the subject of conflicting decisions in the UK as to whether it amounted to publication, and since 2013 legislation has provided protection to a website operator. As noted, protection for search engine operators over “organic search engine results” is pending in Australia.

The complex course of the *Trkulja v Google Inc LLC*¹⁰³ litigation over search engine results is too complicated to recount in detail.¹⁰⁴ The case concerned search engine results that produced images and snippets that the claimant pleaded imputed that he was “a hardened and serious criminal” and “a significant figure in the Melbourne underworld”. Google applied to strike out the proceeding but the primary judge refused to do so, finding that it was strongly arguable that Google’s intentional participation in the communication of the results to search engine users made it a publisher of them. McDonald J also found that the results were *capable* of being defamatory. The Victorian Court of Appeal ventured into other issues about whether Google was a primary or secondary publisher, but the High Court found that this was an error when Google had yet to plead a defence of innocent dissemination.

¹⁰² *Gatley* at [3.034].

¹⁰³ (2018) 263 CLR 149; [2018] HCA 25 (“*Trkulja*”).

¹⁰⁴ For a helpful summary see P. George *Social Media and the Law* at [4.123]-[4.128].

The High Court reversed the Court of Appeal's finding that the claimant had no prospect of establishing any of the meanings. It found that they were *capable* of being conveyed. It would be a matter for trial to decide if an ordinary reasonable user who inputted the terms into the search engine would probably perceive a disconnect between the images and the search terms. It was arguable that someone who had done a search for members of the Melbourne criminal underworld "would rationally suppose" that the response connected to the search.¹⁰⁵ The most obvious, logical connection was that the persons whose images or names appeared in the response under the search headings, or at least some of them, are criminals or members of the Melbourne underworld.

Hyperlinks

Hyperlinks are ubiquitous on social media. *Defteros* is authority that Google was not liable for publication of *The Age* article to which its search result was hyperlinked. It is not authority for the very different proposition that hyperlinks may not be relied upon for *context* in a different factual setting where the reader may be encouraged to read the hyperlinked material and interpret the matter complained of in that context. That context may confirm, qualify or contradict the matter complained of and so is relevant to whether the defamatory meaning alleged was conveyed by that matter.

Unlike the situation in *Defteros*, which concerned a search engine, other online social media publications will be found to adopt, approve, ratify or otherwise publish hyperlinked material. The post may entice and encourage the reader to click on the hyperlink. The defendant's own words may be innocuous in isolation but defamatory because, in the circumstances, they amount to publication of the defamatory hyperlinked material.

Hockey illustrates the possibility of suing over a social media post in isolation, based on the reasonable assumption that at least some readers will not read on by accessing the embedded link. Yet, such an approach calls into question settled rules about not being able to sue over headlines in isolation. Those rules adopt the assumption that the ordinary, reasonable reader reads on. We know that many readers do not, but the law, based on a hypothetical reader, assumes that they do.

Qualified privilege and public interest defences

Following the 2021 amendments to the 2005 Act, a claimant must prove more than the publication of defamatory matter to someone other than the claimant. They must surmount the new "serious harm" threshold. If they do, or are likely to do so, any defendant's attention turns to substantive defences. These include truth, fair reporting, fair comment and qualified privilege. The availability of those defences depends upon many issues, including whether the publisher was actuated by malice.

The availability of public interest defences, by which I mean common law qualified privilege, the category of common law qualified privilege recognised in *Lange v ABC*, statutory qualified privilege and fair comment defences, often will turn on whether the defendant is an originator of the defamatory content or a media entity that publishes the defamatory statements of another. A simple example is where a newspaper neutrally reports that the Leader of the Opposition has accused a government Minister or an official of corruption. A media outlet may not believe in

¹⁰⁵ *Trkulja* at [60].

the truth of that allegation, and may even believe it to be untrue, but act reasonably and in the public interest in facilitating public discussion on that topic. Depending upon the circumstances, the media defendant may have a statutory qualified privilege defence or a *Lange* defence.

In the realm of traditional media, the law developed principles governing the hosting of third-party content, such as letters to the editor. The influential decision of Hunt J in *Hawke v Tamworth Newspapers Co Ltd*,¹⁰⁶ articulated the importance of freedom of expression. The protection afforded to newspaper publishers under the defence of fair comment to publish unpopular or even prejudiced opinions by correspondents is reflected in Lord Diplock's charge to a jury in *Silkin v Beaverbrook Newspapers Ltd*.¹⁰⁷ In Queensland, the 1889 Defamation Griffith Code provided ample protection for newspapers to publish the defamatory statements of third parties. *Pervan v The North Queensland Newspaper Co Ltd*¹⁰⁸ is an example. In *Dorfler v Dwyer & Longbeach Publications Pty Ltd*,¹⁰⁹ the newspaper editor and newspaper successfully defended publishing a letter that accused the plaintiff of being the kind of person who wanted to turn Surfers Paradise into a "'Paradise' for perverts and criminals". At that jury trial, the brilliant and charming editor for whom I acted gave a compelling explanation of her decision to publish such an uncompromising, harsh opinion from a reader.

In the right circumstances, qualified privilege and fair comment defences may apply to defendants neutrally reporting a public figure's statement or "hosting" a letter to the editor. They are harder for a passive publisher that is an intermediary to engage. In theory, the newsagent who sold me *The National Times* in 1983 or the distributor who throws a hard copy of *The Saturday Paper* on my lawn in 2023, might defend any defamatory publication by saying that they believed that I had an interest in receiving information, including harsh and prejudiced opinions by columnists, on matters of legitimate public interest. However, in practice they might struggle to engage defences that require close attention to the decision to publish the actual words that are complained of, rather than the general subject matter of the newspaper. This is why innocent dissemination defences are important to newsagents and other distributors.

For social media publishers, the task of engaging qualified privilege and fair comment defences under Australian law is even harder. In *Google Inc v Duffy*,¹¹⁰ Google failed to engage a defence of qualified privilege for material contained in certain search results. The judgment is a subject of an insightful analysis of the issues of publication and qualified privilege by two brilliant young lawyers who were educated in Queensland.¹¹¹

The problems that confront Google in hosting YouTube videos were brought home to it and others by an award of \$715,000 against it in favour of the former New South Wales Deputy Premier, Mr Barilaro.¹¹² It was common ground in that case that from a certain date Google became liable as publisher for certain "FriendlyJordies" videos uploaded to YouTube by comedian and political commentator, Jordan Shanks. Apart from denying that the matter was defamatory, the only positive defence Google initially maintained was the public interest/qualified privilege defence under s 29A of the *Defamation Act*, in support of which it intended to call the global manager of YouTube's Legal Operations Team. On the first day of trial counsel suggested that the qualified privilege defence under s 29A was irrelevant because Mr

¹⁰⁶ [1983] 1 NSWLR 699.

¹⁰⁷ [1958] 1 WLR 743.

¹⁰⁸ [1991] Aust Torts Reports 81-119; [1991] QSCFC 79.

¹⁰⁹ [1998] QCA 221.

¹¹⁰ [2017] SASCFC 130; (2017) 129 SASR 304.

¹¹¹ M. J. Hafeez-Baig and J. English, 'The liability of search engine operators in defamation: issues relating to publication and qualified privilege' (2017) 24 *Torts Law Journal* 218.

¹¹² *Barilaro v Google LLC* [2022] FCA 650.

Barilaro was not pressing any claim for publication after a certain date. However, when it became apparent that Mr Barilaro had not limited his claim in that way, the s 29A defence was abandoned. Google also abandoned its argument that the matters complained of did not convey the defamatory imputations.

Justice Rares noted Google's stated policy was against hate speech, cyber-bullying and harassment while allowing for artistic and comedic content. He found, however, that Google did not apply its own policies because it did nothing to prevent Mr Shanks' harassment of Mr Barilaro in the videos he published as part of his campaign. I should add that such a finding would have defeated a statutory qualified privilege defence under s 30 that would have required Google to prove that its conduct was reasonable, and a s 29A public interest defence because, in the circumstances, any subjective belief by Google that the publication was in the public interest possibly was likely to be unreasonable.

Litigation and social media

Defamation litigation over social media is a new battlefield. One difficult area is preliminary processes by which an aggrieved individual attempts to ascertain the true identity of the person who has posted defamatory content.¹¹³ Recommendation 6 of the pending reforms address the balancing process.

If an aggrieved party knows the identity of the originator, then a choice has to be made whether to sue the originator, the social media entity that hosted and communicated the defamatory content, or both. For example, Mr Barilaro sued Mr Shanks and reached a settlement with him. He then proceeded against Google.

An interesting area is the selection by a plaintiff of standalone publications, such as tweets, that do not communicate the complexity of the publication they promote and convey a different and more serious defamatory imputation. This was the outcome in *Hockey v Fairfax Media Publications Pty Ltd*.¹¹⁴

There has been a change in profile of defamation cases. The slow death of legacy media, the sacking of journalists, the hollowing out of investigative reporting teams, and the transition of quality newspapers to celebrity puff pieces, sees fewer defamation cases being launched against newspapers and broadcasters. There are celebrated and high-profile exceptions: Ben Roberts-Smith, Rebel Wilson, the Wagner Brothers, Geoffrey Rush, to name a few. However, as a proportion of all defamation cases, cases against mainstream media are in decline.

In 2018, the Centre for Media Transition at the University of Technology, Sydney surveyed defamation cases over a five-year period. Of 189 cases with substantive decisions, 51.3 per cent were digital cases, only 21 per cent of the plaintiffs could be considered public figures, and only 25.9 per cent of the defendant publishers were media companies. This kind of profile can be observed in the Defamation List of the District Court of New South Wales.

Before the internet, defamation litigation could be broadly divided into mass media cases and others about publications with limited circulations. The latter included newsletters and magazines within particular industries, defamatory letters written by a professional or a business about a competitor, nasty things said at P&C meetings, and a variety of what were referred to as "backyard defamations". If a parent defamed a teacher amongst other parents, it was usually something said at the school gate. If the defamed teacher ever found out about it,

¹¹³ See P. George *Social Media and the Law* at [7.5]-[7.40].

¹¹⁴ [2015] FCA 652.

they might have trouble proving exactly what was said. Social media means that similar defamatory publications are posted on Facebook or some other platform with a greater potential for litigation. Judge Muir (as Muir J then was) tried a teacher versus parents defamation case.¹¹⁵ It was a long and complex trial in the District Court, which proved very costly for all the parties.

Often those defamed on social media perceive that they have been defamed “to the world” and fear that anyone in the world can read the defamatory matter on a social media platform. This heightens their hurt and anxiety. Yet, the post may have been read by only a few individuals, and remembered by even fewer. To the extent that limited publications on Facebook pages, such as those hosted by a community organisation, are communicated to factionalised groups, the defamation is likely to be either preaching to the converted or a sermon that falls on deaf ears.¹¹⁶

Defamation law and practice in England responded to cases that involve disproportionate costs by developing the *Jameel* principle under which a court may permanently stay a proceeding as an abuse of process.¹¹⁷ A proceeding may be stayed where the resources involved in the prosecution of the claim (by the parties and the court) are out of all proportion to the interests at stake. That principle has been adopted by some Australian judges in social media cases such as *Bleyer v Google Inc*,¹¹⁸ *Ghosh v Google Pty Ltd*¹¹⁹ and a number of decisions of Judge Gibson, who is in charge of the Defamation List in the District Court of New South Wales. The *Jameel* principle has not found as much favour in Queensland. In one case, Judge McGill SC rejected its application.¹²⁰ In *Watney v Kencian*,¹²¹ the Court of Appeal discussed the principle but it was unnecessary to decide whether it applied.

The recent enactment of a serious harm threshold may filter out some trivial social media cases. The requirement to surmount a serious harm threshold is a welcome addition to the law. However, it will not necessarily filter out trivial cases and, therefore, the defence of triviality might have been retained as a backup protection. It succeeded, for example, in what has been described as the Dennis Denuto case.¹²² If the same publication had been conveyed by social media, the triviality defence would have had similar work to do at the time. Now, if a fairly trivial case is not filtered by the “serious harm” threshold, there is no triviality defence to save a defendant.

Some cases are far from trivial, even if they have a limited publication and a short shelf life. Some texts,¹²³ Judge Gibson’s *Australian Defamation Law and Practice Bulletin*, and the *Gazette of Law and Journalism* record awards in social media and other cases. In Queensland, most defamation cases are tried in the District Court and there have been some very substantial awards in relation to serious defamatory imputations of criminality. For example, in June a District Court Judge awarded \$279,000 in damages to a couple who were defamed by a Facebook user who made a post in a community group suggesting the couple were paedophiles.¹²⁴

¹¹⁵ *Brose v Baluskas* [2020] QDC 15.

¹¹⁶ A description given by Connolly J to a case about letters that were sent by an MP to a number of recipients about a police officer: *O’Shea v Everingham* [1985] QSC 472.

¹¹⁷ *Jameel v Dow Jones* [2005] EWCA 75; [2005] QB 946.

¹¹⁸ [2014] NSWSC 897.

¹¹⁹ [2013] NSWDC 146.

¹²⁰ *Smith v Licht* [2014] QDC 502.

¹²¹ [2017] QCA 116; [2018] 1 Qd R 407 at [48]-[64].

¹²² *Smith v Lucht* [2016] QCA 267.

¹²³ See P. George *Social Media and the Law* at [4.339]-[4.355]; P. George *Defamation Law in Australia* at 634-636.

¹²⁴ *Rodgers v Gooding* [2023] QDC 115.

The difficult question of who is the publisher?

I return to the threshold question of who is legally responsible for a publication? This arises in many different legal contexts. In some contexts, the term “publisher” is not used. If, however, the term “publisher” is a convenient label to describe a party who is to be held legally responsible for the communication of matter via social media, then one needs to be specific about the legal context and what they must do in that context to be liable for the relevant communication. It may be in the context of a criminal law against the dissemination of child pornography, a consumer law against the communication of misleading and deceptive information, the communication of copyright material, the transmission of data in breach of data protection laws, hate speech, words that harass or threaten, or defamatory matter. It may be possible to imagine a social media entity that is a publisher for all or only some of those purposes.

The common law of defamation has struggled to apply traditional definitions of publication to social media. Does the social media entity facilitate and intentionally participate in an act of publication so as to attract legal liability from the time of publication? If not, does allowing third-party content to remain on its platform after being on notice of defamatory content amount to publication by omission?

Are innocent dissemination defences, as modified by an uncertain federal law, fit for purpose for social media operators? How can legal rules encourage moderation of content, at least after a social media host is made aware of defamatory content, while avoiding excessive censorship?

Encouraging moderation - The USA v The rest of the world

If Google’s YouTube platform learned anything from the Barilaro case, it is that social media is not a law-free zone. The law of defamation applies to entities like it that host third-party content, at least once they become aware of it. Australian governments have been struggling with the need to amend Australian defamation laws to deal with social media. The problems have existed for more than two decades. Remarkably, in 2023 we are still trying to work out the effects on defamation law of a federal law that was enacted in 1999.

The High Court in *Voller* clarified the concept of publication. It was not asked to address the application of innocent dissemination defences or the operation of federal law to the facts of that case. Australian governments have decided to supplement existing innocent dissemination defences with a new category of defence for certain internet intermediaries.

Two decades since the rise of Facebook, the American legal system relies upon social media platforms like Facebook and Twitter to moderate themselves.

European nations, the UK and Australia confront issues of online safety, hatred and vilification. Democracies also grapple with what can be done by way of regulatory response to contentious political commentary which, depending upon one’s point of view, could be classified as disinformation or misinformation. Defamation fits somewhere into this regulatory mix.

Laws of the kind developed in Europe and proposed for Australia seek to inject some transparency into the policies and practices that are adopted by social media platforms to monitor and moderate content.

In the modern era of social media it is untenable to treat social media platforms as somehow analogous with libraries, bookshops or a community noticeboard. The 1990s idea that the internet was a new electronic frontier, that would enhance democracy by removing barriers to informed public discussion and create a beautiful marketplace of ideas in which truth would prevail, was noble. Laws that were enacted to protect volunteer moderators of online community discussion forums facilitated the rise of social media giants, like Facebook/Meta. Despite its assertions about being all about a community, Meta is about making money.

Protected by s 230 of the US Act, big social media platforms are largely a law unto themselves. They exert enormous power and may not be greatly influenced by the laws that a nation like Australia enacts. This is not a reason not to enact those laws. Australia's legislation against Big Tobacco is an example of what can be done by one brave nation breaking new regulatory ground. Being realistic, however, there is a limit to what Australia can do to alter the business practices of global social media platforms. Like European democracies, it can pass laws to enhance transparency over digital platforms' policies and practices. It can encourage platforms to abide by codes of practice.

There is a robust discussion in the Academy about moderation¹²⁵ and the need to fix s 230.¹²⁶

Public law plays an interesting role in this context. One contested area in the United States is the application of the First Amendment to social media websites. Subject to narrow exceptions, the First Amendment effectively prevents government agencies and courts from ordering private entities, like newspapers, to publish or to not material. Freedom of speech means that, subject to narrow limitations, media entities enjoy editorial control over what they publish and what they remove. Public spaces, including town halls, are different. So too are public officials. Public forums are supposed to permit the communications that the First Amendment protects. As the result of an innovative lawsuit, Donald Trump was prevented from taking down material from his official site.¹²⁷

The Court of Appeals for the Second Circuit affirmed a Federal District Court's ruling that President Trump's practice of blocking critics from his Twitter account violated the First Amendment.¹²⁸ On July 31, 2020, the Knight Institute filed a second suit in Federal Court against President Trump and his staff for continuing to block critics from the @realDonaldTrump Twitter account.¹²⁹ After the election, President Biden became the petitioning party. The Supreme Court vacated the judgment and remanded it to the Second Circuit to render the case moot.¹³⁰

¹²⁵ E. Douek, 'Content Moderation as Systems Thinking' (2022) 136(2) *Harvard Law Review* 526; N. Minow and M. Minow, 'Social Media Companies Should Pursue Serious Self-Supervision – Soon: response to Professors Douek and Kadri' (2023) 136 *Harvard Law Review Forum* 428; E. Douek, 'The Siren Call of Content Moderation Formalism' in L. Bollinger and G. Stone (eds), *New Technologies of Communication and the First Amendment: The Internet, Social Media and Censorship* (forthcoming); See also the Stanford Law School Podcast 'Moderated Content' <https://law.stanford.edu/directory/evelyn-douek/moderated-content/>.

¹²⁶ D. Citron, 'How To Fix Section 230' (2023) *Boston University Law Review* (forthcoming) https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4054906.

¹²⁷ For a podcast that explains the background to the litigation, see the Knight First Amendment Institute Podcast 'Views on First' (13 January 2023) <https://knightcolumbia.org/content/views-on-first#:~:text=Over%20the%20course%20of%20five,collides%20with%20the%20First%20Amendment>. *Knight First Amendment Institute at Columbia University v Trump* 928 F.3d 226 (2nd Cir. 2019).

¹²⁸ *Knight First Amendment Institute, 'Knight Institute v. Trump'*, <https://knightcolumbia.org/cases/knight-institute-v-trump-2>.

¹³⁰ *Biden v. Knight First Amendment Institute at Columbia University* 141 S.Ct. 1220 (2021); 593 US ____ (2021): https://www.supremecourt.gov/opinions/20pdf/20-197_5ie6.pdf; and see Knight First Amendment Institute, 'Supreme Court Ends Long-Running Lawsuit Over Trump's Now-Defunct Twitter Account', (5 April 2021) <https://knightcolumbia.org/content/supreme-court-ends-long-running-lawsuit-over-trumps-now-defunct-twitter-account>.

In his concurrence, Justice Thomas noted that “applying old doctrines to new digital platforms is rarely straightforward” and that the claimants “have a point ... that some aspects of Mr. Trump’s account resemble a constitutionally protected public forum”. It also seemed “odd to say that something is a government forum when a private company has unrestricted authority to do away with it”. Justice Thomas referred to the control exercised by the social media companies:

“Today’s digital platforms provide avenues for historically unprecedented amounts of speech, including speech by government actors. Also unprecedented, however, is the concentrated control of so much speech in the hands of a few private parties. We will soon have no choice but to address how our legal doctrines apply to highly concentrated, privately owned information infrastructure such as digital platforms.”

Another interesting, new legal frontier arises in what is described as “jawboning”. Jawboning refers to the use of informal methods by government actors to pressure private companies into doing something that the government lacks the legal power to compel the company to do. In the context of freedom of expression, it refers to the tendency of government officials to use informal means, rather than laws, to pressure social media companies to take down what the officials consider to be harmful or offensive speech.¹³¹

Recently, a District Court Judge in the United States ruled that the Biden administration, including the FBI and the Department of Health and Human Services, cannot request social media companies to remove content, as it is protected free speech. Subject to ill-defined exceptions, the ruling bars the administration from contacting social media companies about the content of their sites. The injunction came in a lawsuit led by the Republican attorneys-general of Missouri and Louisiana, who alleged the Biden administration created a “federal censorship enterprise” to curtail posts it viewed as “disinformation”.

A right-wing challenge to self-moderation

At a time when countries like Australia are attempting in a variety of ways to encourage social media platforms to improve their moderation of third-party content, some American States are passing laws that purport to stop social media platforms from moderating their content.

You will recall that during his presidency, Donald Trump declared war on social media companies that moderated his content. Some of them banned him from their platforms. As President, Trump took aim at s 230 of the 1996 Act. He signed an executive order in May 2020, attempting to limit the legal protections of social media platforms.

Both sides of politics in the US seem unhappy with s 230. But like a fractured family, they are unhappy in different ways. Democrats call for more content moderation against hate speech and misinformation. Republicans take a different view. They see social media platforms as

¹³¹ G. Lakier *Informal Government Coercion and The Problem of "Jawboning"*, *The Lawfare Institute* (26 July 2021) <https://www.lawfaremedia.org/article/informal-government-coercion-and-problem-jawboning>; For a recent conversation between Professors Lakier and Douek see ‘Government Platform Communication, Jawboning, and the First Amendment’, *Moderated Content* (10 July 2023) <https://law.stanford.edu/podcasts/governmentplatform-communication-jawboning-and-the-first-amendment/>.

entities that suppress conservative voices. They treat moderation practices as a form of disinformation. In some ways, Republicans want to apply First Amendment values to social media sites, when those social media sites are operated by private entities, not State agencies.

Texas and Florida have each enacted laws. The Texas law makes it unlawful for social media platforms to “censor a user, a user’s expression or a user’s ability to receive the expression of another person based on ... the viewpoint represented in the expression”. Florida’s law aims to prevent a social media platform from wilfully de-platforming a candidate for office. It provides that a social media platform “may not apply or use post-prioritization or shadow banning algorithms for content and material posted by or about ... a candidate”. In addition, platforms are prohibited from removing anything posted by a “journalistic enterprise” based on its content.

Unsurprisingly, the State legislation is under legal challenge from social media platforms and their allies.

The Texas case has gone to the Fifth Circuit. It found the Texas law to be constitutional because it was said to neither compel nor obstruct the social media platforms’ “own speech” in any way.¹³² Echoing the Republican Party point of view, it asserted that the platforms want to eliminate speech rather than protect or promote it.

By contrast, the Eleventh Circuit identified the core question as whether social media platforms, like Facebook and X/Twitter, are “private actors” with First Amendment rights that were constitutionally protected, including their expressive activity in moderating and curating the content that they disseminate on their platforms.¹³³ It ruled that the Florida law interfered with social media platforms’ own speech. It treated the social media platforms as being in the business of disseminating curated collections of speech.

The Fifth Circuit Court, however, treated the social media platforms as “common carriers”. Just as States can impose non-discriminatory requirements on telephone companies, the Court said that the State can regulate conduct in a way that requires private entities to host, transmit, or otherwise facilitate speech. The Eleventh Circuit took the opposite view. It observed that social media platforms are not “dumb pipes”. They are more than servers and hard drives storing and hosting information. They moderate and curate blogs and other information. They are not service providers simply transmitting data from one point to another.

The Eleventh Circuit Court observed Florida’s legislation restricting moderators is so broad that it would prohibit a child-friendly platform, like YouTube Kids, from removing, or even adding an age gate, to soft-core pornography hosted by a Porn Hub which, remarkably, qualifies as a “journalistic enterprise” because it posts more than 100 hours of video and has more than 100 viewers per year.

The two circuit courts express diametrically opposite views about s 230. The Fifth Circuit says that s 230 reflects Congress’ judgment that these platforms are not acting as speakers or publishers when they host user-submitted content. That legislative judgment about the function of online platforms was said to point against finding that the platforms “published” and, therefore, speak the content that their users post. It concluded that s 230 reinforced the conclusion that the social media platforms’ conduct in moderating and removing speech, which might be described as censorship, is not protected speech under the First Amendment.

¹³² *NetChoice LLC v Paxton* 49 F.4th 439 (5th Cir. 2022).

¹³³ *NetChoice LLC v Attorney General, Florida* 34 F.4th 1196 (11th Cir. 2022).

The Eleventh Circuit, on the other hand, concluded that s 230 “explicitly protects internet companies’ ability to restrict access to a plethora of material that they might consider ‘objectionable.’” Section 230 was said to recognise and protect the ability of social media platforms to discriminate among messages by disseminating some, but not others. That was said to be strong evidence that they are not common carriers with diminished First Amendment rights.

In simple terms, two camps in the United States have fundamentally different views about what social media platforms do, and whether they engage in too much or too little moderation. Republicans, particularly Trump supporters, see them as part of a media elite that censored their candidate for President and suppress conservative voices about the 2020 election and many other issues. Two appellate courts considering similar legislation have entirely different perceptions about the social media platforms. They cannot even agree what s 230 of the 1996 Act means.

These cases are heading to the Supreme Court which will be required to consider s 230 of the *Communications Decency Act*. Earlier this year, the Supreme Court of the United States heard argument about s 230 in two related cases.¹³⁴ Some thought that the litigation provided an occasion for it to clarify s 230’s purpose. However, the Court did not have to decide the s 230 issue because, in essence, it found that the social media platforms did not assist or aid terrorism. On that occasion, the Court had abundant material and the benefit of many briefs about the operation of social media platforms. Those who observed the oral hearings remarked that Justice Ketanji Brown Jackson seemed to be the only judge who had done her homework.

For an overseas observer of the American judicial system, the current litigation arising from Texas and Florida has the kind of morbid fascination of an accident that you can see unfolding before your eyes. You want to look away, but you cannot. The scene when these cases reach the Supreme Court of the United States will not be pretty.

Only in the United States could self-moderation by social media giants be seen by a large section of the population and by some senior members of its judiciary as a bad thing. Some legislators pass and some courts uphold laws that prohibit self-moderation.

Conclusion

Reforms to Australian defamation laws that are in the pipeline enjoy bipartisan and national support.

New laws governing liability for internet intermediaries that host defamatory third-party content are to be welcomed. They are long overdue. Proposal 3 builds upon the law of innocent dissemination that was created by common law courts. It creates a new innocent dissemination defence for certain internet intermediaries. It recognises that internet intermediaries often will be ignorant of the content they host until they receive a complaint. It gives the content host 14 days to respond. The content host or other internet intermediary can decide within that period to moderate or remove the offensive content. If it decides not to, then it has a potential exposure to liability.

¹³⁴ *Twitter Inc. v Taamneh et al.* 598 US 471 (2023): https://www.supremecourt.gov/opinions/22pdf/21-1496_d18f.pdf; *Gonzalez et al. v Google LLC* 598 US 617 (2023): https://www.supremecourt.gov/opinions/22pdf/21-1333_6j7a.pdf.

The threat of civil liability is an incentive to the content host to remove clearly defamatory matter, leaving the defamed to pursue the originator of the harmful content for the damage that is done prior to its removal by the content host.

The new scheme's resolution of the rights of the defamed and the rights of third-party content hosts may not be perfect. Some defamed individuals will be aggrieved by the damage done by the content host in allowing the material to be posted and to be downloaded for up to 14 days after complaint. Some whistle-blowers and publicly-spirited citizens who post important, true, but defamatory, statements may be aggrieved that the threat of civil liability prompts a wealthy social media platform to censor their words by removing them from an electronic, public noticeboard.

This imperfect legislative compromise between the rights of the harmed and freedom of expression builds on a common law instinct to not shoot the messenger, at least not before the messenger knows what is in the message and has an opportunity to do something.

Australian law's use of defences for old and new disseminators is to be preferred to the defamation law free zone the US law permits for social media platforms. American law has equipped social media giants with an enormous power to defame and misinform. There is no civil liability incentive to moderate.

Sadly, American judges seem unable to agree what those social media platforms are – conduits of communication or entities that curate and moderate content with First Amendment rights. Surely it is obvious that these entities exercise the power to curate and moderate content, even if the law does not oblige them to do so. Curating and prioritising content is their business. As entities that make those kinds of editorial decisions, they have rights. The law has lagged in imposing legal responsibilities on them. This includes the responsibility, after notice of complaint, to not host indefensibly harmful content.
