DEFENCE OF INNOCENT DISSEMINATION: IS IT AN UNETHICAL SAFE HOUSE FOR SOCIAL MEDIA TO CIRCUMVENT DEFAMATION LIABILITIES?

Authored by
Darshan Rajendren
LL.B (Hons), GDLP,
Adv.Dip (Intellectual Property Law), Dip (Legal Std & Business), Cert (CST)

DEFENCE OF INNOCENT DISSEMINATION: IS IT AN UNETHICAL SAFE HOUSE FOR SOCIAL MEDIA TO CIRCUMVENT DEFAMATION LIABILITIES?

ABSTRACT

The level of ease, accessibility and uninhibited participation facilitated by the modern day digital platforms, especially social media such as Facebook, Instagram, YouTube and Twitter, have introduced a casual and instantaneous approach to communication. Evolvement of internet and the influence of social media encourage the people to say whatever they want about whoever they want, publish it across the world instantly, name it as 'free speech' and continue with their routine life. The ease, reach and anonymity promised often by the social media to its users as internet intermediaries who facilitate publications have led to legal risks, foremostly defamation. Although these social media trends have significantly transformed communications, their impact on the legal principles governing them is less clear and the area of defamation law remains grey with major complexities and inconsistencies. There are substantial and growing debates and concerns on the issue of whether internet intermediaries, such as social media can be held liable as publishers of defamatory matters and if the defences available to them function as an escape route to eliminate their joint legal liability. One such area of concern is the defence of innocent dissemination available to internet intermediaries that is being exploited by social media to evade defamation liability and this paper is centrally structured around this defence and demonstrates the pressing need for reforms in this area.

I. INTRODUCTION

An accusation, which damages a reputation unjustifiably or inexcusably, and published to at least one person other than the plaintiff is considered defamation. The law of defamation in Australia is partly statutory and partly common law, governed by uniform defamation legislation in each state. In Victoria, it is known as the *Defamation Act 2005 (Vic)*. The age-old tort of defamation has applied to all kinds of communication from the time of printing press to the modern social media. Innocent dissemination has been a defence[1] commonly relied upon by the intermediaries. The pro-defendant nature of this defence, due to its

[1] Section 32 of the *Defamation Act 2005* (Vic)

inability to adapt and reflect technological changes, and the very broad definition of "Publisher" in defamation law have made social media rely on this defence of innocent dissemination heavily as an escape route. In recent times, majority of the online defamation cases have involved the defence of innocent dissemination[2]. The general definition of Publisher as per common law is a person who publishes or participates in the publication of defamatory matter and may be liable for defamation, and the participation will require proof of knowledge, awareness, editorial control and capacity over that material and the process.[3] In contrast, a subordinate distributor is a person who is not the primary publisher and has no knowledge, awareness, editorial control or capacity over the publication and the process, hence not liable for defamation. This has been expressly defined in Section 32 of the *Defamation Act 2005* (Vic). Therefore, digital media have been benefitting significantly out of this innocent dissemination defence by pleading and proving their role in publications as subordinate distributors, and not primary publishers. However, the landmark case of *Voller*[4] happened in 2009 and was considered a game-changer, and a barricade in the escape route of social media. The decision of the High Court in this case made operators of social media pages potentially liable for defamation for comments made by third parties on their posts. Considering this case decision as the trigger point, this paper argues that the defence of innocent dissemination functions as an unethical safe-harbour for social media to circumvent defamation liabilities. This paper also strongly recommends an amendment to the innocent dissemination defence to create a default position that social media platforms are primary publishers, given their technological, administrative and financial capacity, editorial control and real-time monitoring facilities. Hence, social media should be liable for defamation jointly with the author or originator of the defaming material.

The recent comments made at the parliament by the Australian Prime Minister Scott Morrison and the Communications Minister Paul Fletcher regarding the pressing needs of social media to be considered liable for defamatory contents have also reaffirmed the need and importance of this research to rethink defamation liabilities of social media and address any escape clauses available to them within the law. The Australian Prime Minister lambasted social media as a "cowards palace", strongly suggesting that these platforms

[2] *Trkulja v Google Inc LLC (No 2)* [2010] VSC 490; *Trkulja v Google Inc LLC (No 5)* [2012] VSC 533; *Tamiz v Google Inc* [2013] 1 WLR 2151; *Google LLC v Duffy* [2017] SASCFC 130; *Defteros v Google LLC* [2020] VSC 219; *Metropolitan International Schools Ltd v Designtechnica Corp* [2011] 1 WLR 1743.
[3] *Dank v Whittaker (No 1)* [2013] NSWSC 1062; *Habib v Radio 2UE Sydney Pty Ltd* [2009] NSWCA 231; *John Fairfax Publications Pty Ltd v Obeid* (2005) 64 NSWLR 485; *Flood v Times Newspaper Ltd* [2012] 2 AC 273; *Lord McAlpine v Bercow* [2013] EWHC 1342; *Cf Bleyer v Google Inc* [2014] NSWDC 897.
[4] *Voller v Nationwide News Pty Limited* [2019] NSWSC 766.

should be treated as publishers when defamatory contents are posted.[5] The Communications Minister Paul Fletcher said that the *Voller*[6] case ruling did not touch on whether Facebook could also be held liable for defamation under Australian law and this question would be examined by the current review of the defamation laws.[7]

The regulation of the defamation content and the liability surrounding it have been areas subject to constant debates and reviews. With the rise and widespread of the internet, especially in the last decade, internet intermediary platforms as a medium have significantly transformed communication with its borderless nature, casual approach, ease, accessibility and anonymity. Internet intermediaries denotes a group of corporations known as telecommunication providers (such as Telstra, Optus etc), internet service providers (TPG, iiNet etc), search engines (Google, Bing etc), content hosts (Youtube, Vimeo etc), e-commerce & digital payment providers (Shopify, ebay, gumtree, Ezypay, PayPal etc) and social media (Facebook, Twitter, Instagram etc). While the research paper acknowledges the benefits of these digital platforms, the liability of internet intermediaries, as a whole, is broader, grey and complex. This research entirely focuses on the role of social media platforms as internet intermediaries in the process of defamation. With the pervasiveness of internet communications and its quotidian nature in contemporary society, comes the need for greater regulatory responses to mitigate online defamation by Imposing liabilities on online intermediaries jointly and severally. However, the function, role and participation of social media are not yet well understood by the legislature. Therefore, the legal responsibility of internet intermediaries for harm is hotly contested and extremely messy.[8]

For defamation liability to be imposed, the legislation should clearly identify the defendant, whether it is the actual author only or the platform that facilitates publication jointly. A prominent justification for imposing liability on social media for defamatory imputations is the need for correction or rectification of harm at instances where the defamatory matter published and the harmed caused by it were wrongfully caused by a third party. As social media disseminates the defamatory matter to the masses, it is effectively enhancing its availability, thereby intensifying and aggravating the harm caused to the reputation of the

---

[5] Byron Kaye, 'A Coward's Palace': Australian PM Slams Social Media Amid Defamation Law Controversy', *Reuters* (Web Page, 7 October 2021) <https://www.reuters.com/technology/australian-law-chief-wants-defamation-rules-fixed-internet-age-letter-2021-10-07/>.

[6] Voller (n 4).

[7] Rod McGuirk, 'Australian Defamation Review to Examine Facebook Liability', *AP News* (Web Page, 6 October 2021) <https://apnews.com/article/technology-business-australia-media-laws-a75ae0c19e7c7caba671d8a4c37fac16>.

[8] Kylie Pappalardo and Nicholas Suzor, 'The Liability of Australian Online Intermediaries' (2018) 40(4) *Sydney Law Review* 469, 475.

victim. This has resulted in majority of arguments turning in favour of imposing liability on social media for defamatory content. Also, with the extension of liability to internet intermediaries based on their knowledge, control and technological expertise over the publications, the Courts have cast doubts on social media claims that they are not responsible for defamatory content on publications facilitated by them. As a result, the social media have constantly tried to plead and prove that they should be protected by the "innocent dissemination" defence[9]. Hence, this defence has been an area subject to exploitation to circumvent defamation liabilities often. This defence is a defence to liability rather than a denial of the publication elements. Another purpose of this research is to demonstrate that the defence of innocent dissemination under statute[10], is articulated in a way that is not 'technology specific'[11] and to invite reconsideration and refinement of the defence of innocent dissemination in defamation law, specifically, to consider social media as primary publishers of all publications facilitated by them.

II. BACKGROUND & LITERATURE REVIEW

Almost 16 years ago, as a response to each State and Territory agreeing to introduce uniform defamation laws, the Legislative Council of those each respective States introduced *Defamation Act*[12] in substantially similar form (The Uniform Act).[13] In 2004 this *Uniform Defamation legislation* implemented the Model Provisions formulated by the former Standing Committee of Attorney-General.[14] As a response to this Uniform Legislation, the *Victorian Defamation Act*[15] was introduced with an aim to provide effective access to defamation related rights. After the Dow Jones v Gutnick decision in 2002, introduction of the Defamation Act in 2005 is considered a major development in the area of defamation law. Parallel to these developments happened the rising of social media giants such as *Facebook* in 2004, *Twitter* in 2006 and *Instagram* in 2010.

[9] Section 32 of the *Defamation Act 2005* (Vic).
[10] Section 32 of the *Defamation Act 2005* (Vic).
[11] David Rolph, 'Publication, Innocent Dissemination and the Internet After Dow Jones & Co Inc v Gutnick' (27 October 2010) *UNSW Law Journal*, 578.
[12] *Defamation Act 2005* (Vic); *Defamation Act 2005* (NSW); *Defamation Act 2006* (NT); *Defamation Act 2005* (QLD); *Defamation Act 2005* (SA); *Defamation Act 2005* (Tas); *Defamation Act 2005* (WA).
[13] 'Defamatory Content Online: The Responsibility of Online Intermediaries' (August 2020) *Birchgrove Legal* <https://birchgrovelegal.com.au/wp-content/uploads/2020/08/OnlineDefamatoryContent.pdf>.
[14] Standing Committee of Attorney-General, *Model Defamation Provisions* (March 2005).
[15] *Defamation Act 2005* (Vic).

Though the statute was created as an effort to regulate and control defamation content, some provisions and aspects of it are not sufficiently flexible to accommodate internet technologies. Thus, the level of internet intermediary liability, specifically social media, is unclear and the struggle to control online defamation has continued. One such statutory variant, which frequently tends to be improper or impractical when applied to online defamation is the defence of Innocent Dissemination. The inflexible characteristic of this concept to accommodate the current internet trends has in return made this defence a mere 'safe house' for digital platforms, especially social media platforms, to evade liability. Therefore, this provision[16] functions as an exploitable legal loophole and the internet intermediaries are highly encouraged to repeat similar defamatory facilitations or even re-offend. As Rolph explains, the defence of innocent dissemination is intertwined with the principles of publication[17] and publication must be proved as a matter of fact and is an essential element of the cause of action in defamation, together with identification and defamatory meaning[18].

Significant consideration has been given to the article by David Rolph[19], which demonstrates the issues in some of the key concepts and ways in which courts have tried to respond to the evolving internet technologies and the challenges it needs to overcome. It raises questions as to whether rules and principles should be technology neutral or technology specific, given the nature of the common law's resistance to adapt to digital changes. The article further analyses the impacts of the *Dow Jones* case[20] on defamation law and the developments as a result in an eight-year period to suggest that, while internet technologies have revolutionised communications, their reflection on defamation law has not been equally radical. The article strongly recommends that certain areas of defamation law need refinement and amendment.

In another separate research[21], the 'deeper pockets' concept was brough to discussion in support of liability imposition on online intermediaries. Suing an online intermediary who has deeper pockets helps not only to restore the reputation of a plaintiff but also receive a larger sum of damages than they would if they were suing an individual. This will reaffirm a society's need to be protected from the consequences of unethical defamation. The article

[16] Section 32 of the *Defamation Act 2005* (Vic).
[17] David Rolph, *Defamation Law* (Lawbook Co, 2016) 139.
[18] *Sims v Jooste (No 2)* [2016] WASCA 83.
[19] Rolph (n 11) 562-580.
[20] *Dow Jones & Co Inc v Gutnick* (2002) 210 CLR 575.
[21] 'Defamatory Content Online: The Responsibility of Online Intermediaries? A Comparative Analysis of Australia, The United States, The European Union, And Canada's Regulatory Responses' (2020), *Birchgrove Legal: Insights Series* (Web Page) <https://birchgrovelegal.com.au/wp-content/uploads/2020/08/OnlineDefamatoryContent.pdf>.

further illustrates that, given the extremely fast spreading nature of contents online, imposing liability can be the key to mitigate reputational harm and protect civil rights of affected parties. It also tries to contextualize the legal discourse behind the urgency to make online intermediaries liable for contents facilitated by them and acknowledges the attempts of the Courts to adapt to the changing nature of online defamation, despite the inability of the statute.

In an important submission by Michael Douglas to the Defamation Working Party, it is said that liability should focus on publication, not authorship[22], based on *Lee v Wilson* decision[23] and the failure to prevent dissemination of a defaming content authored by another person may result in liability[24]. As the innocent dissemination defence is intertwined with the publication rules, the submission establishes that online intermediaries may be considered liable as publishers on application of basic principles.[25] The approach is based on the intermediaries' standard of knowledge and comprehension of the process involved in the publication. The submission constantly emphasises on the divergent views among common law and legislation when considering the level of knowledge of digital platforms such as Google when deciding liability and the unsettled nature of it, strongly supporting my view of the unclear nature of liability imposition.

Another significant discussion paper[26] triggers the interest on the *Section 32* defence and the eligibility of digital platforms under this defence to establish that they are subordinate distributors only, not publishers. The discussion further establishes uncertainties in determining the constructive knowledge of the defamatory content in the context of the intermediary's role in the publication. Recent case decisions have adopted a strict liability test when considering the online intermediary liability, by considering their constructive knowledge. However, it is uncertain as to what constitutes a social media's constructive knowledge and my research also aims to address this question in-depth by discussing the technology-specific accuracy & capacity, availability of fact-checkers, community standards and various other administration & inspection methods of social media.

Therefore, this paper supports my research to demonstrate the unclear or uncertain nature of innocent dissemination defence in general. It explains why this area of defence needs to

[22] Michael Douglas, 'Reforming Defamation Law for The Digital Era: Submission to The Defamation Working Party' (30 April 2019), *NSW Department of Justice,* 11.
[23] *Lee v Wilson* (1934) 51 CLR 276, 287.
[24] *Byrne v Deane* [1937] 1 KB 818.
[25] Matthew Collins, *Collins on Defamation* (Oxford University Press, 2014) 82; *Google Inc v Trkulja* (2016) 342 ALR 504.
[26] 'Discussion Paper: Attorneys-General Review of Model Defamation Provisions – Stage 2', *NSW Government*, 55-59.

be researched further and attempts to clarify the innocent dissemination defence through five questions *(a) Should the defence be amended to provide that the digital platforms are, by default, secondary or subordinate distributors[27]? (b) When can this default position be rebutted? (c) Should a new standalone innocent dissemination defence specifically tailored to internet intermediaries be adopted? (d) How to clarify knowledge or constructive knowledge of internet intermediaries publishing defamatory content? and (e) Are there any other ways in which this defence could be clarified?*

The technology and authority of social media platforms like Facebook and Instagram have shifted their role from mere passive facilitators to potential publishers as suggested by David Rolph[28]. Social media platforms are under no obligation to monitor the content they host, but their community standards and policies suggests otherwise and this is evident from their trends of authority involving actions of content censorship and publication restriction.

Another article by S.H. Chan[29], published by the University of Hong Kong discusses the nature of knowledge possessed, or would have been acquired upon taking reasonable care, which suffices to exclude intermediaries from relying on the defence of innocent dissemination. Though the precedent discussed is not binding in Australia, it is important we compare and contrast the approaches in other jurisdictions due to the borderless nature of social media. Extensively throughout his article Chan has critically analysed the judgement in *Chau Hoi v SEEC Media[30]*, where His Lordship distinguished three types of knowledge possessed by intermediaries or publishers during the publication of a defaming content: (a) Type A knowledge is the knowledge of the gist of content; (b) Type B knowledge is Type A knowledge plus the knowledge that the content is defamatory; and (c) Type C knowledge is Type B knowledge plus the knowledge that there is no valid defence against an action for defamation thereupon.[31] These categories represent crucial differences in knowledge for one to be entitled or disentitled to the innocent dissemination defence. The article further explores on the types of information that make up these different types of knowledge and the severity of liability that changes along with it. Despite the divergent views, this article's approach towards the defence through different categories of knowledge have broadened the scope of the 'knowledge' element of my research and proves to be area of concern.

[27] Section 32(2) of the *Defamation Act 2005* (Vic).
[28] David Rolph, 'Defamation by Social Media' (2013), *Precedent (Sydney, N.S.W)* 16-21.
[29] S.H Chan, 'Innocent Dissemination: The Type of Knowledge Concerned in Shen, Solina Hiolly v. SEEC Media Group Limited' (2016) 25 *Nottingham Law Journal* 152.
[30] *Chau Hoi Shuen, Solina Holly v SEEC Media Group Limited* [2015] HKEC 2703.
[31] Ibid.

A research[32] on the digital defamation trends investigates three key questions of who commences and proceeds with defamation action, who are the publishers these actions are brough against and what are the platforms on which defamatory matter is said to be published. This research provides us with very valuable statistics to reaffirm that our research area is significantly important to be looked out. It found that of the total defamation cases in 2017, 53.3% were digital defamation cases and this adds weight to my research to demonstrate that the defence of innocent dissemination is exploited, resulting in a 200% increase rate of digital defamation cases within a period of 10 years between 2007 to 2017.

III. DEFAMATION

It is essential to understand the characterisation of defamation law and the purpose behind the creation of such legislation to explore the effectiveness of defences to be discussed in this research. The Australian Uniform Legislation for defamation purposely lacks a formal and complete definition of defamation. This is due to the legislation operating by reference to the common law tort of defamation and acting merely as a supplementary provision.[33] The need or purpose of a law to govern defamation is to reach a balance between "society's interest in freedom of expression on one hand and the individual's interest in protecting his or her reputation from unwarranted attack on the other."[34] The Arts Law Centre of Australia explains defamation as "communication from one person to at least one other that harms the reputation of the person, where the communicator (the publisher) has no legal defence."[35]

The cause of action for defamation requires the proof of three elements – publication, identification and defamatory meaning. It is also a legal requirement for the plaintiff to identify the defendant and bring the proceedings within 12 months from the date of publication, though extensions are granted in limited circumstances. Publication occurs when the defamatory content actually gets published and it must also have been published to at least one person other than the person claiming to be defamed. Publication considers the form, to whom and how many, where, when and for how long and about whom.

---

[32] Divya Murthy, 'Centre for Media Transition-Digital Defamation: Defendants, Plaintiffs and Platforms' (2018) *University of Technology Sydney*.

[33] Explanatory Note, Model Defamation Provisions 2005, 3.

[34] New South Wales, *Parliamentary Debates*, Legislative Council, 18 October 2005, 1 (Henry Tsang Parliamentary Secretary).

[35] 'Defamation Law – Online Publication', *Arts Law Centre of Australia* (Web Page, 30 June 2015) < https://www.artslaw.com.au/information-sheet/defamation-law-online-publication/ >.

Identification is the practical ability to identify the person who claims to be defamed directly/indirectly through that publication. It is considered defamatory meaning if the material conveys the imputation. The test for these meanings is whether the ordinary reasonable reader would understand the imputations as directed towards the person considered defamed.

The test for defamation as a whole is whether the ordinary reasonable person or reader would tend to think less of the plaintiff by reason of the defamatory imputations conveyed.[36] There are series of defences available for publishing defamatory material such as justification, contextual truth, absolute privilege, public documents, public concern/interest, qualified privilege, honest opinion and innocent dissemination. Damages for economic and non-economic losses are also available as possible remedies for defamation.

A. Voller's Case, an Australian First

This case[37] is considered the first proceeding in Australia concerning the defamation liability of media companies for comments posted on their Facebook pages. In three separate defamation proceedings, the plaintiff, Dylan Voller sued Fairfax Media, the Australian News Channel and Nationwide News for defamatory comments posted on Facebook in reply to articles posted on the Facebook pages of the Sydney Morning Heralds, The Australian, Sky News, The Bolt Report and The Centralian Advocate during July 2016 to June 2017. Justice Rothman handed down an interlocutory judgement in this proceeding, finding that the three defendant media companies were liable for the publication of comments made by third parties on Facebook pages owned and administered by them. Recently this decision was upheld by the High Court with an absolute majority.[38] The decision also confirms the exposure of social media pages to defamation liability irrespective of intention or knowledge. Their Honours applied the rule in *Webb v Bloch* strictly, stating: "*Webb v Bloch is to be understood to say that a person who has been instrumental in, or contributes to any extent to, the publication of defamatory matter is a publisher. All that is required is a voluntary act of participation in its communication*."[39] This case unveiled an important question as raised by

[36] *Radio 2UE Sydney Pty Ltd v Chesterton* (2009)238 CLR 460.
[37] Voller (n 4).
[38] *Fairfax Media Publications Pty Ltd v Dylan Voller; Nationwide News Pty Limited v Dylan Voller; Australian News Channel Pty Ltd v Dylan Voller* [2021] HCA 27.
[39] Ibid at [32].

the Australian Prime Minister and the Communication Minister: why shouldn't the social media platforms be jointly liable for defamatory content posted?

B. Intermediaries

The role of an intermediary in defamation proceedings is crucial when deciding their entitlement to rely upon the innocent dissemination defences available in common law and statute. Once an intermediary is notified of the defaming content, but chooses not to prevent its publication further, it will be hard to establish such defence as the intermediary will naturally be disentitled from relying on it. However, in Australia, any person who might be described as in 'any degree accessory' to the publication is liable in defamation and includes all participants in and enablers of the dissemination of defamatory matter, subject to this defence availability.[40]

Once a damage to reputation is done, it tends to be irreparable and irreversible. As stated in the *Favell* decision, 'the repetition of an allegation which has been made by an authority such as the police can lead ordinary reasonable persons to understand that, in its context, the person accused is guilty. Allegations or rumours, if repeated, where they involve criminal conduct or not, need to be carefully stated because of the tendency for people to understand and believe that where there is smoke, there is fire.'[41] Intermediaries, especially Social media tend to be the tool that facilitates this repetition of publication, which keeps the 'smoke' (defaming content) alive, regardless of the existence or non-existence of the 'fire' (being true or untrue).

C. Social Media Intermediaries as 'Publishers'

At common law, the definition of publisher in defamation law is too broad and unclear. Any person who participates in a publication in any degree may be liable. In the case of multiple participants, each is jointly and severally liable for defamation.[42] It is a matter for evidence in each case, concerning the level of 'participation' and 'control' attributable to the defendant in

[40] *Trkulja v Google Inc LLC (No 2)* [2010] VSC 490 at 23-26.
[41] *Favell v Queensland Newspaper Pty Ltd* (2005) 221 ALR 186; [2005] HCA 52 at 14.
[42] *Webb v Bloch* (1928) 41 CLR 331 at 363-5; 2 ALJR 282; *Habib v Radio 2UE Sydney Pty Ltd* [2009] NSWCA 231

the publication process.[43] In addition, the person who repeats or republishes defamatory content published by the author or another person may be liable as if the person was the original publisher.[44] To be considered a publisher, the plaintiff must establish that the defendant consented to, or approved of, or adopted, or promoted, or in some way ratified, or in one way or another accepted responsibility for the continued publication of that statement (once it came to the defendant's knowledge).[45] The knowledge and awareness requirement for defamation liability in the context of social media have been constantly debated due to its large scale capacity in terms of technology, administration and finance, which by default equips them with the required knowledge and awareness of materials posted in their platforms. A perfect example to this is Facebook's Content Filtering system which publishes high quality posts and comments first.[46] Regardless of the usage of an algorithm or an administrative effort, if social media platforms are capable of deciding the quality of its users' expressions or opinions, it is understood to have involved adequate knowledge and awareness of all of its publications.

D. Social Media Vs Other Media

There may be questions on the significance and uniqueness of social media when compared to the traditional media such as TV and radio, and the international media. The radio and TV have been broadcasting to large audiences for decades and the power and reach of international media have also expanded in the recent years enormously. But social media is distinguishable from them due to the real-time engagement of its users, its complete control over posts and the ability to publish them in multiple online platforms instantly. The uniqueness of social media lies in the simple and casual approach by which a material can be posted online by its users. Another significant feature of social media is that the news and materials need not be edited or arranged in a formal way before publication, and can be posted in its original and authentic form.

---

[43] *Thompson v Australian Capital Television Pty Ltd* [1996] HCA 38; (1996) 186 CLR 574; *Google LLC v Duffy* [2017] SASCFC 130 per Kourakis CJ, at [92].
[44] *John Fairfax Publications Pty Ltd v Obeid* (2005) 64 NSWLR 485; *Flood v Times Newspapers Ltd* [2012] 2 AC 273; *Lord McAlpine v Bercow* [2013] EWHC 1342.
[45] *Urbanchich v Drummoyne Municipal Council* (1988) A Def R 50-035; *Bishop v New South Wales* [2000] NSWSC 1042.
[46] 'News Feed FYI: Showing More Quality Content', *Meta for Business* (Web Page, 23 August 2013) < https://www.facebook.com/business/news/News-Feed-FYI-Showing-More-High-Quality-Content>.

IV. COMPETING INTERESTS

While this paper invites reforms and refinement of *Section 32*[47] defence provision, one might argue that this will lead to overregulation and over censorship of expressions, which will have a chilling effect on free speech. Therefore, it is important to address this potential argument first to understand the crux of the research.

An academic analysis[48] demonstrates that Australian defamation law fails adequately to take into account the interests of the readers/recipients of publication when it deals with defamation defences and also supports some of the views of competing interests discussed in this paper. The author reaffirms that the entire purpose of the defamation law is to strike a balance between reputation and freedom of expression and these align with the respective positions of the publisher and the victim, but do not engage directly with the interests of the recipients. This literature strongly recommends that it is vital for defamation defences to be reformed and emphasises the importance of rethinking the suitability of innocent dissemination defence to balance and uphold both dignity and free speech.

A. Reputation

The plaintiff's reputation is the primary interest in defamation law. However, it also tries to strike a balance between right to reputation and freedom of speech.[49] Lord Denning defined Reputation as "what other people think of a person". He stated, *"a man's character, it is sometimes said, is what he is in fact is, whereas his reputation is what other people think he is."*[50] There are three competing conceptions of reputation inherent to defamation law: reputation as honour, reputation as property and reputation as dignity.[51] Reputation as dignity is the most valued and recognisable concept in the present defamation law, as it tries to affirm the dignity of a person is respected at all times and to reinforce the rules of civility.

---

[47] *Defamation Act 2005 (Vic).*
[48] Michael Gillooly, *Third Man: Reform of the Australasian Defamation Defences* (Federation Press, 1st ed, 2004).
[49] *Silkin v Beaverbrook Newspapers Ltd* [1958] 1 WLR 743; *Chakravarti v Advertiser Newspapers Ltd* (1998) 193 CLR 519; *Ballina Shire Council v Ringland* (1994) 33 NSWLR 680.
[50] *Plato Films Ltd v Speidel* [1961] AC 1090.
[51] Robert Post, 'The Social Foundations of Defamation Law: Reputation and the Constitution' (1986) 74 *California Law Review* 693.

B. Free Speech

Freedom of speech is recognised as an important value at many levels and it should be protected as it enhances the quality of democracy, autonomy and promotes truth and morality, which in turn is good for humanity[52]. While the research acknowledges the importance of free speech, whether reputation should be upheld and protected at the expense of free speech is beyond the research scope of this paper. This research further ensures that the suggestions to be made for reforms to the defence of innocent dissemination will not undermine free speech, but instead, will enhance the quality and merit of it and prevents its misuse. "*There is a general rule that a man, while exercising freedom of speech, must take care not to defame his neighbour*".[53] Therefore, reforms in this area will assist in upholding the rights of reputation and free speech.

V. SOCIAL MEDIA POLICIES & STANDARDS

There are 4.20 billion active social media users worldwide and an average of 2 million are joining them every day. In the last 12 months, global active social media users increased by 490 million, which is an increase by 13.2%.[54] In Australia, as of January 2021, there were 20.50 million social media users with a penetration rate of 80% of the total population.[55] The power of social media is commendable, as well as threatening due to its unique integration tools such as likes and reactions, follow buttons, photo sharing, live streaming etc, and the integration of social media with mobile phones, which makes it accessible any time. As this paper demonstrates that social media should be made liable for publications facilitated by them and the innocent dissemination defence should not be made available to rely upon, it is necessary to understand the type of commitment and policy measures created by social media to mitigate defamation. Through these commitments, standards and policies the technical capacity and capabilities can be determined to establish their control and knowledge in publications facilitated by them.

---

[52] Des Butler and Sharon Rodrick, *Australian Media Law* (Thomson Reuters, 5th ed, 2015) 8.
[53] *New South Wales Parliamentary Debates*, 2nd Series, vol 35 (1909) 3103.
[54] 'Facts & Figures // Social Media Statistics for 2022', *Social Media Perth* (Web Page, 08 September 2020) < https://www.smperth.com/resources/social-media-statistics/>.
[55] Ibid.

A. Facebook Community Standards

Facebook's revenue amounted to roughly 86 billion US dollars in 2020.[56] However, it's service for more than two billion of its users have continued to be for free, allowing them to freely express themselves across countries and cultures, in many languages. One of Its key role is to keep abuse and defamatory contents off the service. It further recognises that the platform creates new and increased opportunities for abusive and degrading contents, and tries to limit these expressions by giving importance to authenticity, safety, privacy and dignity. These precautions are carried out with the help of experts in fields such as technology, public safety & human rights, independent third-party fact-checkers, academics, other organisations and various signals from machine learning model including community feedback.[57]

B. Instagram Community Standards

To ensure the dignity and respect of the community, Instagram removes content that contains credible threats or hate speech and contents that target private individuals to degrade or shame them. Similar to Facebook, Instagram also uses technology, community feedbacks and third-party fact-checkers to assist in this process. Importantly, Instagram works with third-party fact-checkers across the globe who review content in over 60 languages and are certified through the non-partisan international Fact-checking Network to help identify, review and label publications that go against these standards.[58]

VI. INNOCENT DISSEMINATION

Firstly, this vital part of the research will establish that social media should be considered as primary distributors or publishers with complete editorial control, technical capacity,

---

[56] 'Facebook: Annual Revenue 2009-2020', *Statista Research Department* (Web Page, 5 February 2021) <https://www.statista.com/statistics/268604/annual-revenue-of-facebook/>.
[57] 'Facebook Community Standards', *Transparency Center: Facebook* (Web Page) <https://transparency.fb.com/en-gb/policies/community-standards/>.
[58] 'Reducing the spread of false information', *Help Centre: Instagram* (Web Page) <https://help.instagram.com/1735798276553028/?helpref=hc_fnav>.

awareness and knowledge of the content published in their platforms by its users, and any lack of knowledge over its own content must foremostly be due to negligence only. Therefore, social media should be considered as primary publishers of its content and be liable for any defaming imputations, thereby disentitling them from the defence of innocent dissemination. Secondly, it will also demonstrate how this defence works as a safe house for social media to evade defamation liability.

Section 32 of the *Uniform Defamation Legislation*[59] provides a defence of innocent dissemination, which protects a 'subordinate distributor' from liability. A subordinate distributer is defined in sub-section 32(2) as a publisher who is (a) not the first or primary distributor of the matter, (b) 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 before it was first published. Without limiting this definition, sub-section 32(3) includes a list of persons that are not the primary distributors of matter. This includes who participate in the capacity of a bookseller, librarian, wholesaler or retailer such as news agents, printers and postal services. It also includes 'an operator of, or a provider of access to, a communications system by means of which the matter is transmitted, or made available, by another person over whom the operator has no effective control'.[60]

A defendant should not only prove to be a subordinate distributor, but also that they did not know, nor ought reasonably to have known that the matter was defamatory[61] and this lack of knowledge was not due to any negligence on the part of the defendant[62]. These 2 provisions make it clear that once a subordinate publisher is notified on the defamatory content, it cannot rely on this defence of innocent dissemination and risks losing the statutory benefits of it. In *Tamiz v Google Inc*[63], the English Court decided that the defendant became a publisher of the material as it had been notified of the presence of that material, thereby making itself responsible for the continued presence of that material. However, in Australia, the stand taken in *Trkulja*[64] judgement is significant, as it confirms any person, including all participants, in 'any degree accessory' to the publication is liable in defamation. *Dow Jones & Co Inc v Gutnick*[65] remains the leading Australian authority on internet defamation, which gives significant consideration to publication and innocent dissemination. This case raised a range of questions, most importantly: what constitutes publication, should technology

[59] *Defamation Act 2005* (Vic); *Defamation Act 2005* (NSW); *Defamation Act 2006* (NT); *Defamation Act 2005* (QLD); *Defamation Act 2005* (SA); *Defamation Act 2005* (Tas); *Defamation Act 2005* (WA).
[60] Section 32(3) of the *Defamation Act 2005* (Vic).
[61] Section 32(1)(b) of the *Defamation Act 2005* (Vic).
[62] Section 32(1)(c) of the *Defamation Act 2005* (Vic).
[63] *Tamiz v Google Inc* [2013] 1 WLR 2151.
[64] *Trkulja v Google* (No 5) [2012] VSC 533.
[65] *Dow Jones & Co Inc v Gutnick* (2002) 210 CLR 575.

specific rules be developed to cover internet publications or can the existing principles be adapted and is the response to these changes by courts and legislature sufficient to counter the challenges. This case suggests that, while the internet has made revolution in communications, their impact on defamation law are not equally radical and remains unclear.[66]

A. Publication Capacity

As this paper attempts, a significant aspect of it is to explain and establish that social media cannot be considered as defendants who publish defamatory content merely in the capacity of subordinate distributors due to a number of key factors that lead to inconsistencies with what the legislation defines. Innocent dissemination defence is available to social media given that the defendant published the defamatory matter merely in the capacity of a subordinate distributor. To be entitled to this defence social media should prove that it is neither the primary distributor nor the author or originator and did not have any capacity to exercise editorial control over the content before it was published.[67] Social media often tries to prove that it has no effective control over the person who is the author of the defaming matter or over the content published, in order to entitle themselves for this defence.

As principles of publication and innocent dissemination are intertwined, publication remains the foremost factor in deciding the liability. In *Byrne v Deane*[68] the plaintiff held the defendant liable for defamation, not on the basis that they had composed the defaming matter, but on the basis that the defendant facilitated the continuous presence of it, thereby becoming its publisher.

Sub-clause 32(2)(b)[69] defines a subordinate distributor as a person who is not the author or originator of the matter. Taking a divergent view, Justice Dixon stated that liability focuses on publication, not authorship.[70] Accordingly, even a subordinate distributor may be considered liable for publication. Examples of subordinate distributors mentioned in the legislation are booksellers, news agents, librarians, wholesalers or retailers, broadcasters, communication system provider and persons who print/produce or distribute on the instruction of another

[66] David Rolph, 'Publication, Innocent Dissemination and The Internet After Dow Jones & Co Inc v Gutnick', *University of New South Wales Law Journal* (2010) 33(2), 563.
[67] Section 32(2) of the *Defamation Act 2005* (Vic).
[68] *Byrne v Deane* [1937] 1 KB 818.
[69] *Defamation Act 2005* (Vic).
[70] *Lee v Wilson* (1934) 51 CLR 276, 287.

person.[71] The nature and control of these subordinate distributors when compared with the modern day social media are highly dissimilar and discordant. A fine example to demonstrate the capacity of social media is the ban imposed on the former U.S President Donald Trump from their platforms indefinitely, as a precautionary measure against potential harmful content to be shared.[72] This proves the decisive role social media play when it comes to which posts, from who and about whom to be published in their platforms. A mere facilitator will not possess the capacity to make publication decisions based on a content.

Another strong proof to demonstrate the capacity of social media is the 27-pages *Content Moderation & Censorship Guidelines*[73] revealed by Facebook in 2018. It encompassed dozens of topics including hate speech, misrepresentation, disinformation etc, and the company's censor's called 'content moderators' and third-party fact-checkers ensure these guidelines are applied to all its publications.[74] This capacity of social media to exercise instant moderation and censorship over material already published or to be published in its platform itself shows that these intermediaries play a vital role than just being mere facilitators of publications. Importantly, social media can never fit into the subordinate distributors category such as booksellers, newsagents, postal services etc, given in Sub-section 32(3)[75].

Afterall, even if one argues that social media play a mere passive role in facilitating the publication and dissemination of it, failure to prevent dissemination of defamatory material authored by another person itself may still result in liability[76] and this failure could be due to various different reasonings such as negligence and underperformance relating to technology, administration and editorial.

---

[71] Section 32(3) (a)-(h) of the *Defamation Act 2005* (Vic).
[72] Dipayan Ghosh, 'Are We Entering a New Era of Social Media Regulation?', *Harvard Business Review* (Web Page, 14 January 2014)
<https://hbr.org/2021/01/are-we-entering-a-new-era-of-social-media-regulation>
[73] Julia Wong & Olivia Solon, 'Facebook Releases Content Moderation Guidelines – Rules Long Kept Secret', *The Guardian* (Web Page, 25 April 2018)
<https://www.theguardian.com/technology/2018/apr/24/facebook-releases-content-moderation-guidelines-secret-rules>.
[74] 'Facebook Reveals Its Censorship Guidelines for the First Time – 27 Pages of them', *Los Angeles Times* (Web Page, 24 April 2018)
<https://www.latimes.com/business/technology/la-fi-tn-facebook-guidelines-20180424-story.html>.
[75] *Defamation Act 2005* (Vic).
[76] Rolph (n 66) 569-570.

B. Editorial Control

The traditional intermediaries had very limited editorial oversight once published, whereas, modern day social media have instant editorial control including immediate censorship and alteration of material published in their platforms. 'The Content Moderation & Censorship Guidelines' of Facebook is a fine example to show the usage of human editors by social media, rather than just depending on Artificial Intelligence(AI) algorithms. The moderators and third-party fact-checkers employed globally by social media determine whether posts, comments, messages or images violate its policy. According to Law Insider, 'Editorial Control' means the right to review, formulate for, or to exercise a veto over the appearance, text, use or promotion of the product.[77]

In other words, the editorial control of social media is its authority to review, edit, moderate, censor or ban any material published, being published or to be published, to comply with its guidelines and standards. This ability to control contents is only possible if the platforms are continuously supervised. In the judgement of *Thompson v Australian Capital Television*, the High Court held the respondent broadcaster was liable for the defamatory content, 'as it was not merely a conduit because it had the ability to control and supervise the material published'.[78]

A media report in June 2021 states, "In recent weeks, Instagram and Facebook have censored posts focused on COVID-19 in India and protests in Colombia and Palestine – with little explanation as to why".[79] Twitter blocking all links to a New York Post story on the business dealings and related information of Hunter Biden, and the continued ban on Post's account demonstrates Twitter's editorial control[80].

In two significant case decisions of New Zealand, *Wishart v Murray*[81] and *Karam v parker*[82], it was held that the defendants were liable as publishers, *inter alia*, due to their ability to control and moderate contents. However, the Australian common law lacks clarity on control

---

[77] 'Editorial Control Definition', *Law Insider* (Web Page)
< https://www.lawinsider.com/dictionary/editorial-control>.
[78] *Thompson v Australian Capital Television Pty Ltd* [1996] HCA 38.
[79] Rishika Pardikar, 'Social Media Companies Like Instagram Are Censoring Dissent', *Jacobin* (Web Page, 1st June 2021)
<https://www.jacobinmag.com/2021/06/censorship-facebook-instagram-twitter-india-palestine-colombia>.
[80] 'Twitter's Censorship Method: Dorsey Doesn't Seem to Regret His Company's Partisan Favouritism', *The Wall Street Journal* (Web Page, 28 October 2020)
<https://www.wsj.com/articles/twitters-censorship-method-11603927356>.
[81] *Wishart v Murray* [2013] NZHC 540.
[82] *Karam v Parker* [2014] NZHC 737.

and moderation capabilities of defendants and a similar approach is considered essential. Courtney J held the defendants liable as publishers and stated, "the defendant had the ability to control content by deleting posts and blocking individual users from the page[83]; the defendant had control over the content, and actively moderated it"[84]. Content control and moderation capabilities are understood to be the key elements when deciding an intermediary's liability. " Twitter has the Twitter Rules, Reddit has a content policy, Youtube has Community Guideline and Facebook has Community Standards. Facebook has a policy team made up of lawyers, public relations professionals, ex-public policy wonks, and crisis management experts that makes these rules. They are enforced by roughly 7,500 human moderators, according to the company."[85] As evident, the modern day social media have the required capacity to exercise editorial control, inclusive of moderating, deleting and blocking the contents at any stage of the publication facilitated by them. Therefore, it can be convincingly established that social media should not be allowed to avail itself of the defence of innocent dissemination based on any question of capacity to exercise editorial control over its contents, as its technical and administrative capabilities are much efficient and effective when compared to the traditional media.

1. Technical Capability

'It is unclear which types of internet intermediaries would be considered 'subordinate distributors' given that some may be considered to have the technical capacity to exercise editorial control'. [86] This position is taken due to the differences in the nature and capacity of traditional internet intermediaries and the social media. AI algorithms, community feedbacks, Facebook Artificial Intelligence Researchers (FAIR) and various other technological signals (image matching technology, identical content label etc) are used to exercise editorial control over contents that do not comply with its policy standards and guidelines.

Every minute, people upload over 136,000 photos and 293,000+ statuses on Facebook.[87] The AI derive knowledge from the posts' data (comments, likes, shares etc), models, algorithms, applications, and hardware and software infrastructure. Jerome Presenti, Vice

[83] *Wishart* (n 81) at [116].
[84] *Karam* (n 82) at [23].
[85] Jason Koebler and Joseph Cox, 'The Impossible Job: Inside Facebook's Struggle to Moderate Two Billion People", *Motherboard: Tech by Vice* (Web Page) <https://www.vice.com/en/article/xwk9zd/how-facebook-content-moderation-works>.
[86] Discussion Paper (n 18) 53.
[87] Michael Iyam, '7 Revealing Secrets – How Facebook is Using Artificial Intelligence', *ITChronicles* (Web Page) <https://infoforu.org/7-revealing-secrets-how-facebook-is-using-artificial-intelligence/>.

President of AI at Facebook said*, "the rate of AI usage is remarkable at Facebook. Our number of AI engineers and developers is more than doubling every year"*.[88] Through AI, Facebook also tries to solve some of the biggest issues faced across seven major categories such as Hate speech, Terrorism, Nudity, Graphic Violence, Spam, Suicides and Fake Accounts.[89] This powerful technical competency gives social media an immense editorial control over its contents.

2. *Administrative Capability*

It is the availability of qualified personnel or plans to undertake the overall administrative responsibility to ensure high degree of editorial control. Community standards & guidelines, Content moderators, third-party fact-checkers, academics and other organisations are the key players in this administrative process. Instagram 'work with third-party fact-checkers across the globe who review content in over 60 languages and are certified through the non-partisan International Fact-Checking Network to help identify, review and label false information'.[90]

One may argue that identifying defamatory content and exercising editorial control over it would be costly, considering the human resource needed. I believe an increased administrative cost in the private sector is still a safe option than an increase in the public sector costs, by opening the floodgates to large number of claims and lawsuit. In Australia, out of the total defamation cases, 53.3% were digital defamation cases, which was a 200% increase in the rate of digital defamation cases within a period of 10 years between 2007 to 2017.[91] Creating a clear default position that social media are not primary publishers will not only control this floodgate effect, but will also encourage resolution without litigation, to a great extent. As long as this area of defence is grey and unclear, all parties, especially the 'deep pocket' social media will take chances to exploit it.

Importantly, it is no excuse that monitoring & moderating every single post in social media would significantly increase the compliance costs of it. Social media is an integral operational and promotional tool for businesses around the globe, as Facebook alone has

[88] Ibid.
[89] Ibid.
[90] Instagram Community Standards (n 58).
[91] Murthy (n 32).

generated $85.9 billion revenue in 2020[92]. Together with these huge commercial benefits should they be prepared to bear the administrative costs that help mitigate liability and comply with policies. If not, they should be prepared to bear the legal consequences. Understandably, there is difference between facilitating publications of 4 million users and 4 billion users. If one has the capacity to facilitate a medium for that many users, it should also have the capacity to act as a responsible medium.

C. Reasonable Awareness

Publication capacity, control, awareness and knowledge are intertwined and have similar influential factors and issues to be considered multiple times, applying it to each sub-section of the defence. Both sub-sections 32(1)(b) and (c)[93] impliedly refer to awareness of social media with regard to the content posted in their platforms. A defendant must prove that the defamatory content could not be identified by exercising reasonable awareness or it was impossible to be aware of the imputation due to various reasons. In general, these sub-clauses require the defendants to prove that its ignorance was not due to negligence[94], and they were reasonably aware. The main issues to be considered are: (a) how can social media be aware and in which ways can this be done, and (b) accuracy of the defamatory judgement by social media based on that awareness. These issues are still to be settled as there is very limited common law authority in this area and has not been discussed enough. The centrality of social media awareness and the aversion to technology specific rules, are clearly reflected in the statutory defence of innocent dissemination under the uniform defamation legislation.

The capacity and control established by social media enables it to be aware of all material published in its platforms regardless of it being defamatory or not. This is achieved by instantaneous technical and administrative expertise, resources and control available to social media. But if social media use this expertise to be aware of defamatory imputations is a matter of its choice. One can argue that the defendant had lack of knowledge of the defamatory content, despite being reasonably aware. This will raise questions on the

[92] Mansoor Iqbal, 'Facebook Revenue and Usage Statistics (2021)', *Business of Apps* (Web Page, 24 September 2021)
< https://www.businessofapps.com/data/facebook-statistics/>.
[93] *Defamation Act* 2005 (Vic).
[94] Section 32(1)(c) of the *Defamation Act* 2005 (Vic).

accuracy and efficiency of the methods used by social media to acquire awareness. However, the most common issue of social media as understood from many recent cases is its negligence in acquiring awareness or exercising reasonable care when facilitating publications. To counter negligence allegations, the defendant must prove that they did everything reasonably necessary to determine whether the content was defamatory , including the steps taken by the defendant in researching, editing, and fact-checking their work.[95] In general, for negligence, the Courts may consider factors such as amount of research, trustworthiness of sources, attempts to verify questionable views or soliciting views and whether the defendant followed other good journalistic practices.[96] Given the availability of resources at hand for social media to acquire immediate awareness regarding publication involving violence, nudity, sexually suggestive content etc over its platforms and the ability to restrict them simultaneously, the lack of knowledge on defamatory content is to be considered as due to complete negligence only. The privilege to facilitate wide publication among global audience should not undermine its content awareness. While this paper is aware of the practical difficulties in being aware of thousands of publications every minute, it is also not advisable to let these platforms authorise all posts regardless of the unlawful and objectionable imputations they might contain. This undifferentiated and neutral authorisation of all publications posted by its users would be inconsistent with the platform's community standards & guidelines and the role of the administrators specified in them.

D. Knowledge

'Knowledge' has been the most crucial and controversial element, which is the foremost deciding factor in the defence of innocent dissemination for social media. This paper tries to establish that social media by default are to be considered as defendants with adequate knowledge or reasonably expected to have known the content of material published in its platforms. The statute states that a defendant can rely on the defence of innocent dissemination if it can be proven that the defendant neither knew, nor ought reasonably to have known, that the matter was defamatory.[97] As we discuss in this section, common law have had diverging views on whether knowledge of the defamatory imputation is necessary, as established in the *Byrne*[98] case, or whether it is not necessary. Knowledge on defamatory

[95] 'Proving Fault: Actual Malice and Negligence', *Digital Media Law Project* (Web Page, 10 September 2021) <https://www.dmlp.org/legal-guide/proving-fault-actual-malice-and-negligence>.
[96] Ibid.
[97] Section 32(1)(b) of the *Defamation Act 2005* (Vic).
[98] *Byrne v Deane* [1937] 1 KB 818.

matter is not considered necessary as per the nature of tort's strict liability. The reason for diverging views and judgements on cases with similar facts is due to the failure of the statute and common law to understand or define the clear differences between the role of publishers and intermediaries. These discrepancies have increased in the last decade together with the dominance of social media. The focal point of this research is the nature of knowledge possessed or ought reasonably to have been acquired, which would entitle or disentitle social media to this innocent dissemination defence. Are the social media only liable once the defamatory content is brought to its knowledge or attention after notice? If contents with violence, terrorism, adult nudity, sexual activity, self-harm etc are identified and removed instantaneously or rapidly without anyone reporting it, how can social media expect to gain reasonable knowledge on defamatory content only upon notice and not before that? If the argument is based on human input vs algorithms, isn't it a biased or unfair approach? Or even be technical and administrative incompetence amounting to negligence? Social media Community Standards & Guidelines[99] discussed above expressly confirms the instantaneous involvement and control through human inputs and resources over all its publications for compliance. Can't we consider the formation of these guidelines as an effort to obtain compulsory knowledge on all material being facilitated by social media? To know a material complies with the guidelines, one should be aware of its contents completely. How this is achieved, be it through human inputs or automated algorithms, should not be the concern of the Courts when deciding on defamation liabilities. With these views, I now turn to the common law and the unclear stand/approach of the Courts in relation to 'knowledge' and establish how it supports my dogma.

1. Australia

In Australia, two recent case decisions handed down in 2012 and 2014 have been inconsistent and the Courts have had diverging views on the knowledge element which makes them a publisher. In the 2012 case *Trkulja v Google*[100] (Trkulja), Google published search results of "Melbourne criminal underworld photos" and "Melbourne underworld criminals" that included images of Mr Trkulja (defendant) mixed with images of convicted Melbourne criminals, and also texts and predictions that referred and linked him with terms like "is a former hit man", "criminal' and "underworld". In his decision, Beach J held that 'it was open to the jury to include that Google intended to publish, even before notification,

[99] Facebook (n 57); Instagram (n 58).
[100] *Trkulja v Google* (No 5) [2012] VSC 533.

everything produced by its automated systems, which its employees created and allowed to operate, because Google is similar to a newsagent'.[101] His Honour, disagreed with any general rule that passive intermediaries cannot be held liable as publishers, because such a rule would cut across the library and newsagent line of cases[102] and therefore, Google can be considered a publisher that is liable, even without notice of the content of the material[103]. This significant case decision has demonstrated two important views: (a) knowledge of the content is required, regardless of being notified or not, and (b) defendant neither knowing nor ought reasonably to have known the defamatory content is not a defence if it's a publisher.

2. New Zealand

In 2014, just after two years of *Trkulja* decision, *Bleyer v Google Inc*[104] was decided differently, contradicting the views of Beach J. McCallum J in this case was satisfied that there were sufficient evidence to suggest that there were no human inputs in the search process of Google except the algorithms. The judge expressly stated that 'performance of the function of the algorithm is not capable of establishing liability as a publisher at common law, therefore, Google cannot be liable as a publisher of the results produced by its search engine, at least prior to notification'. [105] This case suggests that automated systems with no human input cannot be expected to have knowledge, or ought reasonably to have known the material's content. These two cases discussed clearly show the disagreement and the diverging views of the Judges in relation to the knowledge element. However, *Trkulja* decision has opened doors for the New Zealand common law to take a similar approach and this is evident from the case *A v Google*[106], being consistent with *Trkulja*. In this case, search engines were considered to be publishers, thereby making them liable for any defamatory content.

The New Zealand Courts have provided several significant precedents that shifted from the actual knowledge requirement in *Wishart*[107] and *Karam*[108]. Even if the defendant had no knowledge, it will still be liable, provided it ought, in the circumstances, to know that the

[101] Ibid at [18].
[102] Ibid at [28].
[103] Ibid at [30].
[104] *Bleyer v Google Inc* [2014] NSWSC 987.
[105] Ibid at [83].
[106] *A v Google New Zealand Ltd* [2012] NZHC 2352.
[107] *Wishart* (n 81).
[108] *Karam* (n 82).

postings are being made that are likely to be defamatory. In *Wishart v Murray (Wishart)*[109] the Court considered the possibilities of a Facebook page being held liable for defamatory content posted on the page by other users, as the defendant had the ability to control content by deleting posts and blocking individual users from the page.[110] Courtney J in her decision stated that "those who host Facebook pages or similar are not passive instruments or mere conduits of content posted on their Facebook page. They will be regarded as publishers of defamatory material made by anonymous users in two circumstances. The first is if they know of the defamatory statement and fail to remove it within a reasonable time in circumstances that give rise to an inference that they are taking responsibility for it. A request by the person affected is not necessary. The second is where they do not know of the defamatory posting but ought, in the circumstances, to know that postings are being made that are likely to be defamatory".[111] *Karam v Parker (Karam)*[112] was also decided by the same Judge, Courtney J, where she referred and affirmed the two circumstances discussed. Her Honour held that the defendant, who was a Facebook page administrator, to be a publisher of posts and statements made by third parties[113], including third party comments[114]. *Wishart* and *Karam* strongly suggest that a defendant can be liable even without the knowledge of its content. The liability position taken by Courts on Facebook pages in these two cases should be maintained when considering the liability of social media platforms in Australia. For a better understanding, Facebook pages should be considered as miniature models of the full-sized social media and both should be given the same level of liability weightage and importance, if not more, to social media, given its capacity and influence.

3. Hong Kong

The Hong Kong case law *Chau Hoi Shuen, Solina Holly v SEEC Media Group Limited* (2015) (HKCFA) is significant when discussing 'knowledge', as His Lordship Fok PJ delivered the only reasoned judgement. His lordship distinguished between three types of knowledge: (a) Type A knowledge is the knowledge of the gist of content; (b) Type B knowledge is the Type A knowledge plus the knowledge that the content is defamatory; and (c) Type C knowledge is Type B knowledge plus the knowledge that there is no valid

[109] *Wishart* (n 81).
[110] Ibid at [116].
[111] Ibid at [117].
[112] *Karam* (n 82).
[113] Ibid at [19].
[114] Ibid at [23].

defence against an action for defamation thereupon. For this purpose of innocent dissemination defence, Type C knowledge is not considered, but Type A and B. In the judgement of *Oriental Press Group Ltd*[115] case, these knowledge types have been referred to and was held that a main publisher is liable because of his control and knowledge of the contents of publication (i.e. Type A knowledge); it being irrelevant whether he knows that the content was defamatory (i.e. Type B knowledge).[116] This case strongly suggests that the same test of knowledge must be applied in deciding whether a subordinate publisher can avail of this defence.[117] This case law is significant as it suggests that the control and knowledge of the contents itself would be sufficient to hold a party liable, regardless of them having knowledge of the defamatory imputation.

4. England & Wales

This approach has already been followed by Patteson J in *Day v Bream* (1837)[118], where His Honour stated that a defendant is prima facie liable even if he had no knowledge of the contents of the publication[119] and 184 years after, this should be the default position that should be taken in relation to all publications facilitated by social media. If such a precedent can be set in 1837, only with the understanding of printing technology (as it was also the year in which the first US electric printing press patented by Thomas Davenport[120]), arguing on the knowledge element in this era of hand-held 3D printers only shows the failure and inflexibility of the legal system to accommodate and acknowledge technological development & brilliance.

However, the common law of different jurisdictions are gradually moving towards technology specific approaches to keep up with the modern trends, and the decisions in *Trkulja*[121], *Karam*[122], *Wishart*[123], *Chau Hoi*[124] and *A v Google*[125] clearly demonstrate this shift. As they all suggest, Social media as defendants should be held liable even without knowledge, provided it ought to know the content of all the materials published, given the technical &

---

[115] *Oriental Press Group Ltd v Fevaworks Solutions Ltd* (2013) 16 HKCFAR 366; [2013] HKEC 1025.
[116] Ibid.
[117] S.H Chan (n 29) 153.
[118] *Day v Bream* (1837) 174 ER 212 (Assizes).
[119] Ibid at [56].
[120] 'Historical Events in 1837', *On This Day* (Web Page) <https://www.onthisday.com/events/date/1837>.
[121] *Trkulja* (n 40).
[122] *Karam* (n 82).
[123] *Wishart* (n 81).
[124] *Chau Hoi Shuen, Solina Holly v SEEC Media Group Limited* (2015) (HKCFA).
[125] *A v Google* (n 106).

administrative brilliance they possess and the strict policy compliance exercised by them. A stand as such will shape the future direction of Australian laws with respect to social media defamation. If they want to avoid liability for defamation, then they should take a more active role to acquire reasonable knowledge of materials published in its platforms, rather than trying to evade the knowledge criterion.

VII. UNETHICAL 'SAFE HOUSE'

For the reasons discussed, it is necessary that social media are jointly held liable for third party defamation on its platforms. Different jurisdictions have tried different ways to address issues in this area by being pro-plaintiff or pro-defendant[126] and this is based on multiple factors, including its influence in the local economy and e-commerce market. Social media provide real-time insights into their users' tastes, preferences, trends etc to its business clients across marketing and e-commerce touch points. Social media can even predict which photos will resonate with a brand's unique audience the most, in real time. Tracking and measuring public engagements in real time has not only been vital to the humongous success and revenue-making of social media, but also to its clients who seek various services and benefits, be it governments, political parties, celebrities, multi-national organizations, religious groups etc.

As we are aware, social media such as Facebook and Instagram are also widely criticised for its very limited and biased efforts to curb the spread of disinformation, misinformation, hate speech and election meddling. Political actors use social media to influence elections and political behaviours and trends, including the past US elections.[127] While social media benefits financially from the spread of these incendiary and misinformation-laden posts[128], it has also led to increasing political polarisation in the society[129]. This was corroborated by findings from the US Brennan Centre for Justice. A research team led by journalism and

[126] Russell Weaver "Defamation Law in Turmoil: The Challenges Presented by the Internet" (2000), *JILT-Warwick* (Web Page)
<https://warwick.ac.uk/fac/soc/law/elj/jilt/2000_3/weaver/>.
[127] Robert Bond, et al, 'A 61-Million-Person Experiment in Social Influence and Political Mobilization' (13 September 2012), *NIH Public Access: Author Manuscript.*
[128] Michael Brand, 'Facebook is Tilting the Political Playing Field More Than Ever, And It's No Accident', *The Conversation* (Web Page, 27 October 2020)
<https://theconversation.com/facebook-is-tilting-the-political-playing-field-more-than-ever-and-its-no-accident-148314>.
[129] Casey Newton, 'The Libel Bias on Social Networks Isn't Against Conservatives: It's Towards Polarization', *The Verge* (Web Page, 11 April 2019)
<https://www.theverge.com/interface/2019/4/11/18305407/social-network-conservative-bias-twitter-facebook-ted-cruz>.

communications professor Young Mie Kim identified a range of Facebook troll accounts deliberately disseminating division by targeting the society, with posts to instigate hate, fear, outrage, conflict and hostility, and these troll accounts have been spreading defaming information on personalities of interest not just during election campaign periods, but also prior or during the build-up to it.[130] Therefore, one cannot argue that these political expressions are exempt from any legal actions as they are made only during political campaigns. Even if that is the case, we must take notice of the fact that social media is global and there is some political activity or election happening somewhere around the world all the time. This is a very serious concern considering the amount of misinformation, hate speech and fake news that disseminate overall, globally. 'Facebook has drawn widespread criticism for its failure to remove posts that clearly violate its policies on hate speech including posts by Trump himself'.[131] It is clear that social media have been biased and have favoured sides for various benefits and they use defamation, hate speech, misinformation and various other methods to achieve it. Facebook expressly has created policies that exempts politicians from its fact-checking program and knowingly hosts misleading content from politicians, under its "newsworthiness exception".[132] As stated by Mark Zuckerberg himself, "when left unchecked, people on the platform engage disproportionately with such content".[133] This disproportionality often leads to defaming expressions towards one party from the other and can benefit social media immensely. These coordinated inauthentic behaviour of social media for monetary gains have put its integrity and credibility in jeopardy, and cannot be expected to come with clean hands to defend themselves using the defence of innocent dissemination.

Social media's superpower of real-time technology and capacity to predict its users behaviours, influence the general public, tilt the political fields, make and break governments, make a 'no one' into a celebrity over-night and manipulate the thinking trends of a society have undoubtedly been successful due to the existence of the innocent dissemination defence, an unethical escape route for them to evade defamation liability with ease.

---

[130] Young Mie Kim, 'New Evidence Shows How Russia's Election Interference Has Gotten More Brazen', *Brennan Center for Justice* (Web Page, 5 March 2020) <https://www.brennancenter.org/experts/young-mie-kim>.
[131] Brand (n 128).
[132] Nick Clegg, 'Facebook, Elections and Political Speech', *Meta* (Web Page, 24 September 2019) <https://about.fb.com/news/2019/09/elections-and-political-speech/>.
[133] Brand (n 128).

A. Social Media and Politics

We have seen the significant and influential role media plays in the political fields throughout the history. They have contributed to the change of governments both intentionally and accidentally. However, social media like Facebook have been tilting the political playing fields more than ever, and it is no accident.[134] According to a University of Oxford study[135], 70 countries, including Australia, practised either foreign or domestic election meddling in 2019. This was up from 48 in 2018 and 28 in 2017.[136] The study said Facebook was "the platform of choice" for this. Within this social media platform cyber troops use a variety of communication strategies such as (i) disinformation, (ii) mass reporting, (iii) data-driven strategies, (iv) trolling, doxing or harassment and (v) amplifying content. Most of these strategies involve various forms of defamation such as fake news, memes, videos, trolls, harassments, hate speech etc. The research by the Oxford University has also revealed some shocking statistics on how social media manipulate the voters and build propagandas[137]:

- *87% of countries used human accounts*
- *80% of countries used Bot accounts*
- *11% of countries used Cyborg accounts*
- *7% of countries used hacked or stolen accounts*
- *71% spread pro-government or pro-party propaganda*
- *89% use computational propaganda to attack political opposition*
- *34% spread polarising messages designed to derive divisions within society*
- *75% of countries used disinformation and media manipulation to mislead users*
- *68% of countries use state-sponsored trolling to target political dissidents, the opposition or journalists*
- *73% amplify messages and content by flooding hashtags*

Facebook and Twitter were also accused over US election actions due to their editorial decisions about what to take down, label or leave unaltered, which made them publishers

[134] Brand (n 128).
[135] Samantha Bradshaw & Philip Howard, ' The Global Disinformation Order: 2019 Global Inventory of Organised Social Media Manipulation', *Computational Propaganda Research Project, University of Oxford.*
[136] Brand (n 128).
[137] Ibid 11-15.

rather than just distributors.[138] “Federal law gave you the ability to stand up and grow without being hit by lawsuits, but you have used this power to run amok”, said Republican Senator Blackburn[139]. Republican Senator Lindsey Graham, chairman of the Senate Judiciary Committee, added: "When you have companies that have the power of governments, have more power than traditional media outlets, something has to give."[140] These statistics and political views clearly demonstrate that social media have more power than all other media outlets existed so far, making it unique, influential and domineering.

VIII. RECOMMENDATIONS

This research paper strongly recommends a few measures to counter this defence section from working as a safe house for social media to circumvent defamation liability, via legislative amendments or by establishing a Commission, which will be discussed below. These recommendations will have a fair and balanced approach towards both social media and the defamed party, rather than being pro-plaintiff aimlessly.

A. Legislative Amendments

An amendment to the innocent dissemination defence should be created to introduce a default position that social media are primary distributors. This will clarify the legislative position regarding the controversial involvement of social media in publication. Alternatively, this could be achieved by inserting a provision or paragraph which expressly states that social media are not to be considered as subordinate distributors. Alternatively, a standalone subsection could be added to clause 32, which would exempt or disentitle social media from relying on this section of the defence. This would make the presumption non-rebuttable, thereby making social media act expeditiously on any defamatory material published in its platforms.

This standalone section could also specify a ‘reasonable time period’ within which social media should identify and take actions on a defamatory content. This time period should be

[138] Leo Kelion, ‘Facebook and Twitter Grilled Over US Election Actions’, *BBC News* (Web Page, 17 November 2020)
<https://www.bbc.com/news/education-54974819>.
[139] Ibid.
[140] Ibid.

decided by keeping in mind the capacity, real-time engagement, technical advancement and the reach of social media on one hand, and a comparison to the scale of damage that could be caused to a victim within that time and after that time (grapevine effect), on the other. This will be a fair and well-balanced approach towards both parties involved. Introduction of a section as such would also ensure that social media manage their risk of liability for defamation in a better way and do not attempt to circumvent its liability by exploiting the law.

B. A 'Commission'

An independent 'Commission for social media defamation' that will be vested with authority to review and deliver binding decisions on defamation cases reported by an Australian person should be established. This will divert the ubiquitous defamation suits from the Courts system and take the pressure off it immensely. Social media should develop a tailored 'Report defamation' feature in their platforms, which will directly be brought to the attention of the Commission. This system should provide and promote the parties with non-litigious dispute settlement methods and also, a merits review option if any party wish to challenge the decision of the Commission.

Importantly, this whole system, including the Commission and the staff should be funded by digital platforms, if possible. The Commissioner who will be appointed to this Commission should be a retired judge of the Australian Courts. This recommendation is useful as it will help the Courts system by taking the pressure off it, cutting down the costs and saving time.

IX. CONCLUSION

The laws relating to defamation liability of internet intermediaries are well defined and settled in many countries, especially the United Kingdom and also New Zealand to some extent. However, the current less-settled Australian law sits awkwardly between the two models of traditional intermediaries and internet intermediaries, even three, as social media model should be considered entirely separate. There are clear deficiencies with the innocent dissemination defence that makes it pro-defendant and biased, and needs major reforms and amendments, in the best interest or benefit of the general public.

As recommended in this paper, an amendment providing a default position that social media are primary publishers who cannot avail themselves of this defence will provide greater certainty for the users and general public, knowing that if they are defamed, there is a legal requirement in place to pressurize the social media to expeditiously identify, review and remove/delist the imputation within the time specified, without the need of any complaints or notice procedure, to avoid defamation liability. This default position should be applied automatically after a specific time frame given (such as 24 to 48 hours) and this will work as an immunity time period for social media, within which they can take necessary actions to avoid defamation liability. Imposition of a time limit will give all parties a fair, reasonable and balanced approach to mitigate defamation and resolve the matter. Creation of a Commission will help dispute the matters, on a case-by-case basis, if the parties wish to proceed further and seek remedies. Only the matters with merits should be allowed to challenge the decision of the Commission at the Court and this will reduce the caseload of the Court system drastically.

This paper also demonstrated that the current defence of innocent dissemination is ineffective from a victim's perspective, when dealing with online publications due to social media's broad capacity, technical brilliance, extremely effective control, real-time awareness and in-depth knowledge gaining ability, which are distinct characteristics when compared to traditional intermediaries, both online and offline. Inopportunely, *Section 32*[141] defence has only managed to be an unethical safe house for social media to circumvent or evade defamation liability, despite being an active and influential role-player in the publication. Therefore, it appears necessary to immediately reform the innocent dissemination defence to remove the detriments of it to the general public and to uphold the trust & confidence in the judicial system. Social media defamation is a twenty-first-century problem which remains regulated and highly influenced by the nineteenth-century laws. While this paper acknowledges that social media is a gift, if used in the right way, it is time we rethink and reform the defence of innocent dissemination, to fulfill the community's greater need for protection, in all possible ways and forms.

[141] *Defamation Act* 2005 (Vic).