Liability for Online Anonymous Speech: Comparative and Economic Analyses*

Ronen Perry: Professor, Faculty of Law, University of Haifa.
Tal Z. Zarsky: Associate Professor, Faculty of Law, University of Haifa.

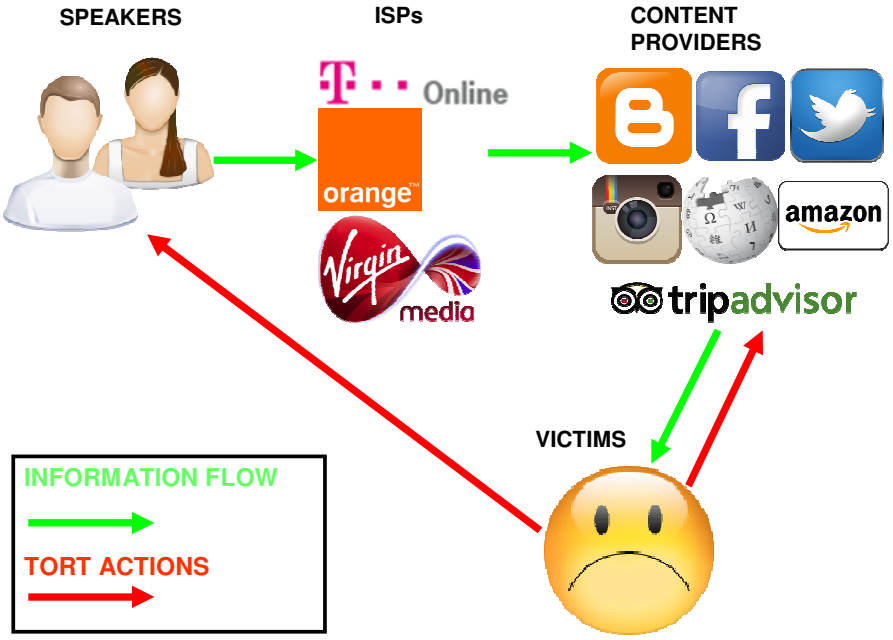
* The authors are grateful to participants in the 13th Annual Conference on European Tort Law (24-26 April 2014, Vienna) and to the anonymous referee for valuable comments on earlier drafts.

I. INTRODUCTION

Web 2.0 enables average people to connect to the internet through Internet Service Providers (ISPs) and contribute content to websites. Although technology is rapidly advancing, and old online platforms constantly make way for new ones, a few contemporary examples may be useful. Simple users can (1) ask questions or provide answers at online forums; (2) write blogs or make comments on others' blog posts; (3) publish customer reviews of travel-related services at TripAdvisor.com or of books at Amazon.com; (4) take part in multi-user discussions on social networking services, such as Facebook, Twitter or LinkedIn; (5) share photos and videos or make comments on others' photos on Instagram or Flickr; and (6) participate in collaborative writing projects, such as Wikipedia.

Some of these user-contributions may be defamatory. Since Web 2.0 has turned every person into a potential public speaker, it has dramatically increased the group of potential wrongdoers; and because almost everyone is using the internet, the audience is much larger than before, increasing the expected impact of every potentially defamatory statement. In this Article we address the special and common case in which a defamatory statement was made by an anonymous or a pseudonymous user on an online platform operated by an identifiable other.[1] By shielding users from the potential impact of their statements on their own interests (for example, reputation or employment), anonymity and pseudonymity encourage people to participate, and arguably to demonstrate less constraint and more insensitivity. Although to some extent these consequences may be warranted (in fueling exchange of ideas and public debate), they also increase by a wide margin the likelihood of defamation. The question we wish to address is simple to state, but very difficult to answer: who should be liable for online anonymous defamation?

[1] But see *T Zarsky/NNG de Andrade*, Regulating Electronic Identity Intermediaries: The 'Soft eID' Conundrum (2013) 74 Ohio State Law Journal 1335 for a discussion of the growing trend of mandating the use of real names by online intermediaries (such as Facebook).



This article's first major contribution lies in recognising that the legal response to online anonymous defamation must be viewed holistically, taking into account at least two components of the liability regime, and the interaction between them: (1) the ability to bring an action against the platform enabling the defamatory statement, which we call 'the content provider,' and (2) the ability to bring an action against the anonymous user, whom we call 'the speaker.' As we demonstrate below, these two components, and the interactions between them, are designed differently in different legal systems.

A claim against the content provider requires acceptance and enforcement of some form of substantive indirect liability. Put differently, to allow such claims, the law must impose liability on intermediaries who did not actively contribute to the wrongful conduct, and possibly had limited actual knowledge of it. Indirect liability raises numerous problems, especially in the context of online speech. Thus, in some jurisdictions content providers' liability is very limited or even excluded.

A claim against the anonymous speaker does not require special adaptation of substantive defamation law, but a procedural apparatus which enables the plaintiff to uncloak the speaker's identity. To enable plaintiffs to do so, the law must first devise a process for ordering the relevant content provider to turn over identification information (usually an IP address) about the speaker.[2] But even after receiving such information, the

[2] Our analysis makes several assumptions. First, it assumes that intermediaries will not voluntarily share speakers' personal information with plaintiffs. This assumption is fair, given privacy law restrictions and intermediaries' interest in maintaining a positive reputation in the market. Second, it assumes that the relevant intermediaries obtained such

plaintiff must reach out to at least one more intermediary. He or she must identify the ISP linked to the relevant IP address (using the WHOIS directory), and then request contact information of the user associated with this IP address.[3] At times, the IP address will point to a public or a multi-user computer. In such cases, the plaintiff must engage further to uncloak the speaker's identity—turning to libraries, workplaces, internet cafés or other realms in which multiple users log on.[4] The legal system might be reluctant to provide plaintiffs with uncloaking abilities. Even when it enables de-anonymisation, speakers may evade liability by using advanced anonymisation tools, such as Tor,[5] by connecting through public hot spots which do not require registration, or when the relevant records are lost along the way.

In this article we examine four general models for handling the problem of online anonymous defamation. We do so first from a comparative perspective, showing that each model has been endorsed, with adaptations and modifications, by some jurisdictions. Next, we evaluate each of these models from a cost-benefit perspective, pointing out their relative strengths and weaknesses. The ultimate goal is to determine which model is economically superior.

Our analysis makes several assumptions, which constitute methodological concessions, for the sake of simplicity and clarity. First and foremost, we assume that the delicate balance between freedom of speech and the right to reputation in any given jurisdiction accurately reflects prevalent values and preferences within this jurisdiction. We do not aim to challenge existing boundaries of liability for defamation, but to investigate which method implements this balance in the most cost-effective way in the context of online anonymous speech.

Second, we assume that online anonymous defamation is a distinct category of online anonymous speech, which justifies special analysis. Some scholars have called for a unified discussion of content providers'

---

information and retained it after the fact. This too is a fair assumption as platforms retain such information for internal use. Furthermore, sometimes retention of user information is required by law. The EU Data Retention Directive is a good example. See *T Verbiest/G Spindler/GM Riccio/A Van der Perre*, Study on the Liability of Internet Intermediaries (2007) 82-83 <http://ec.europa.eu/internal_market/e-commerce/docs/study/liability/final_report_en.pdf>. Recently, however, the ECJ found this Directive invalid, because it infringed the right to privacy and the right to the protection of personal data in a disproportional manner. *Digital Rights Ireland Ltd v Minister for Communications* (ECJ, 8 April 2014) <http://curia.europa.eu/juris/ documents.jsf?num=C-293/12>. Relevant state laws regarding retention, however, still stand.

[3] *N Gleicher*, John Doe Subpoenas: Toward a Consistent Legal Standard (2008) 118 Yale Law Journal 320, 328. ISPs usually obtain and retain such information for their ongoing operations, such as billing.

[4] See *Gleicher* (fn 3) 328 (discussing the two-step process).

[5] Running or using such tools is not illegal. *AM Froomkin*, Anonymity and the Law in the United States, in: I Kerr/C Lucock/V Steeves (eds), Lessons from the Identity Trail: Anonymity, Privacy and Identity in a Networked Society (2009) 460.

liability for users' illegitimate speech,[6] including defamation, copyright violations, misrepresentation of products on sales platforms, speech crimes such as incitement, creating or selling hate speech or obscene content,[7] and perhaps also disseminating computer viruses and the like. We reject this approach not only because other instances of online anonymous speech are subject to special legal rules in many jurisdictions, but also because we believe they raise different problems and concerns.[8]

Third, we assume that an anonymous statement about another person may cause harm to that person's reputation. Indeed, many anonymous speakers lack sufficient reliability and credibility, so their statements are unlikely to cause any harm. But when people continuously use a specific alias on a particular platform, they gain credibility as repeat players. And even one-shotters may cause reputational harm in many instances, as cases decided in various jurisdictions demonstrate.

Fourth, we assume that the candidates for indirect liability are only platforms with some control over user-generated content. This assumption is consistent with applicable law in most jurisdictions. Online intermediaries that function as mere passive conduits, such as ISPs and other telecommunication service providers, including Skype and WhatsApp, are not normally liable for transmitted defamatory statements. Although this legal reality may be contested, the case for liability seems very weak.

Fifth, we assume that all parties are subject to the same jurisdiction. Special problems arise with respect to transnational defamation—posted by foreigners or on foreign platforms.[9] Bringing foreign speakers to local courts may be difficult, and ordering foreign platforms to disclose information about their users may be a very thorny task. Moreover, variance in liability regimes among jurisdictions may channel online activities to jurisdictions with more lenient attitudes to defamation in general, to content providers or to anonymous speakers.[10] Although these problems existed prior to the new media age, the internet has exacerbated them significantly. We ignore them for simplicity, noting that they can be alleviated to some extent by transnational unification and harmonisation endeavors.

---

[6] *A Hamdani*, Who's Liable For Cyberwrongs? (2002) 87 Cornell Law Review 901, 903.

[7] 47 US Code § 230.

[8] With respect to copyright violation see, in the US, See, 17 US Code § 512. In the EU, a similar regime seems to apply. See *A Chander*, How Law Made Silicon Valley (2014) 63 Emory Law Journal 639, 676.

[9] For a discussion of the problems associated with transnational defamation see *L Levi*, The Problem of Trans-National Libel (2012) 60 American Journal of Comparative Law 507.

[10] Questions concerning international jurisdiction and applicable substantive law may also arise and complicate any litigation. See, eg, *F Wagner-von Papp*/*J Fedtke*, Germany, in: European Tort Law Yearbook 2011, 242, 252-255, nos 26-32 (discussing the 'Seven Days in Moscow' case).

II. A COMPARATIVE ANALYSIS

A Overview

The comparative analysis shows that there are four paradigms for combining the two components presented in the introduction (namely, recognising content providers' liability and uncloaking anonymous speakers). The first paradigm, which currently applies in the US, bars content providers' indirect liability, but facilitates identification of the speaker. The second, currently used in Israel, recognises content providers' fault-based liability but does not provide procedural tools for identifying the speaker. The third, which we call the EU framework, enables the victim to request identification of the speaker, but at the same time bring an action against the content provider. Although there is variance among member states, this model seems to comply with the relevant Directives and European courts' decisions. The fourth, which was recently adopted in England, enables the victim to pursue a claim against the speaker, and if the speaker is unavailable, imposes liability on the content provider. We call this 'residual indirect liability.' In this Part we will explain how each of these systems works.

The analysis in this section is admittedly broad-brush. To tackle the major points we set aside important yet relatively minor distinctions. For example, legal theory and practice distinguish between the content provider's liability before and after being notified of a possible speech-related tort.[11] In general, when content providers are considered mere 'intermediaries,' they are not exposed to liability prior to notification. However, when considered 'publishers' liability might attach even without such notification.[12] Other distinctions concern the nature of content providers' business model,[13] the existence of manual and automated forms of filtering that can be used to limit expected harm,[14] and the form of discourse content providers facilitate.[15] These distinctions, while important, would complicate the comparative analysis, rendering it almost futile in

[11] The question as to what form of notification indeed puts the content provider on 'notice' and diminishes its ability to argue it had no knowledge of a possible defamatory statement is complex. Such a notice might include not only an allegation but, possibly, an explanation for the unlawfulness of the relevant content. See *G de Wilde*, Case Law: Tamiz v Google Inc, Google May be a Common Law Publisher (21 February 2013) Inforrm's Blog <http://inforrm.wordpress.com/2013/02/21/case-law-tamiz-v-google-inc-google-may-be-a-common-law-publisher-gervase-de-wilde/>

[12] See *D Voorhoof*, Qualification of News Portal as Publisher of Users' Comment May Have Far-Reaching Consequences for Online Freedom of Expression: Delfi AS v. Estonia (25 October 2013) Strasbourg Observations <http://strasbourgobservers.com/2013/10/25/qualification-of-news-portal-as-publisher-of-users-comment-may-have-far-reaching-consequences-for-online-freedom-of-expression-delfi-as-v-estonia/>

[13] *Delfi AS v Estonia* [2013] ECtHR, Application No 64569/09 (10 October 2013), para 89.

[14] *Id* para 87.

[15] *Id* para 27 (accounting for the fact that comments were integrated into a news portal); see also generally *M Lavi/T Zarsky*, A Contextual Theory for Online Intermediary Liability (2013) 43 Hebrew University Law Review 631 (in Hebrew).

understanding the main picture. We will return to some of them in the economic analysis.

B US model: exclusive direct liability

1 Indirect liability

In the specific context discussed here, the US legal regime mostly blocks lawsuits against online content providers. Most potential legal actions against online intermediating content providers are effectively blocked by sec 230 of the federal Communications Decency Act [16] (hereinafter 'sec 230'). However, American law offers some redress against the anonymous speaker. It allows plaintiffs to bring actions against intermediaries—requiring them to disclose information regarding potential defendants. Such information, in turn, might enable lifting the veil of anonymity and bringing direct legal action against anonymous speakers.

The enactment of sec 230 famously responded to several earlier cases which addressed platform liability under the common law. The common law of defamation distinguishes between three types of intermediaries. 'Common carriers,' such as telephone companies, only transmit information and are not liable for defamation. 'Distributors' of published material, such as bookstore owners, distribute content without control of the content, and are liable only when they have actual knowledge of the defaming nature of the publication or had reason to know. 'Publishers,' such as newspapers, exercise significant control over published content, and are subject to strict liability.[17]

In *Cubby v CompuServe*,[18] the court found that CompuServe, which operated special interest forums, should not be liable for content posted as part of a non-moderated discussion. CompuServe provided its users with access to a daily newsletter about broadcast journalism and journalists. CompuServe itself merely uploaded the information, without reviewing its content. The plaintiffs argued that the newsletter included false and defamatory statements about a competing newsletter they published. The court held that CompuServe was the 'distributor,' rather than the 'publisher' of the information, and therefore not liable.[19] In *Stratton Oakmont, Inc v Prodigy Services Co*,[20] the court found that Prodigy, a bulletin board operator, was liable as a 'publisher' for defamatory statements made by its users about the plaintiff. Such statements, voiced by an unidentified user on the 'Money Talk' computer bulletin board, attributed criminal and fraudulent acts related to securities trading and

[16] 47 US Code § 230.

[17] *DJ Solove/PM Schwartz,* Information Privacy Law (4th edn 2011) 184.

[18] *Cubby, Inc v CompuServe, Inc*, 776 F Supp 135 (SD NY 1991).

[19] See also *Solove/Schwartz* (fn 17) 184.

[20] *Stratton Oakmont, Inc v Prodigy Services Co*, 1995 WL 323710 (NY SCt 1995). For a recent discussion, see *Chander* (fn 8) 651.

financial markets to Stratton Oakmont and its president.[21] Here, the online board was advertised as one that featured editorial control.

The jurisprudential turn seems to have generated an unfortunate outcome. 'Good Samaritan' filtering of digital content by content providers might prove beneficial to potential victims, but also to other online users who would enjoy a friendlier virtual environment. Alas, the joint reading of the *Cubby* and *Stratton Oakmont* decisions created a legal environment in which content providers had a strong *disincentive* to moderate online discourse, because moderating exposed them to the risk of liability. The court in *Stratton Oakmont* acknowledged this problem, but conjectured that market forces would provide proper incentives for content providers to engage in monitoring. This was an interesting yet arguable premise which had only limited empirical support.[22]

Pressures from the internet industry quickly led to the enactment of sec 230, which superceded the *Stratton Oakmont* decision. This law's stated objective was to promote, among others, fora for diverse public discourse,[23] as well as preserve vibrant and free markets online.[24] Specifically, sec 230(c)(1) of the CDA states that online service providers should not be considered as 'publishers' of 'any information provided by another information content provider.'[25]

The change sec 230 brought about was vast. This was quickly made evident in *Zeran v America Online, Inc*[26] which came about the following year. The case featured defamatory statements pertaining to Zeran made on an AOL message board. Specifically, an anonymous user posted messages advertising that Zeran was selling shirts and other materials with slogans glorifying the 1995 Oklahoma City Bombing. These postings also included the plaintiff's home phone number. As a result, Zeran received numerous threatening phone calls. Zeran contacted AOL and requested removal of these messages, and it is disputed as to how long it took AOL to react. This unfortunate turn of events repeated itself several times, and Zeran brought an action against AOL. Zeran did not sue the anonymous user, claiming he was unable to do so because AOL did not maintain records of its users.

AOL was not found liable for the defamatory posts, even though it had relevant knowledge from a certain point, and even though it would have been considered a 'publisher' under traditional defamation law. The court held that the claim was barred by sec 230. Subsequently, sec 230 has

---

[21] The actions of Stratton Oakmont and its president Daniel Porush were recently featured in the Martin Scorsese film 'The Wolf of Wall Street,' which was based on the memoirs of the firm's founder, Jordan Belfort.

[22] One might easily argue that to the contrary, in some instances, firms actually benefit from an aggressive discourse unfolding throughout their platform, as such spirited discussions generate interest and thus additional participation and even advertising-related revenue.

[23] 47 US Code § 230(c)(1).

[24] *Id.*

[25] *Id*.

[26] 129 Federal Reporter (F) 3d 327 (4th Cir 1997).

provided online content providers with effective immunity[27] in a variety of contexts, and from a broad range of causes of action.[28] Such immunity extended to 'publishers' and 'distributors.' One commentator recently summarised this issue by stating that sec 230 'largely immunized online service providers from secondary liability for most torts committed through their services.'[29]

The balance struck by and the broad immunity arising from sec 230 is not without criticism. In view of the law's broad interpretation by the courts, some commentators have called for its amendment and subsequent limitation. They argue that although sec 230 has proven invaluable to the notion of free speech in the internet context, it might be unfair towards individuals who find their reputation tarnished online.[30] Furthermore, sec 230 enables harassment which is highly problematic, especially when the harmful anonymous comments are directed at women and minorities.[31] Yet given sec 230's central role in the regulation of US platforms and its possible contribution to their success,[32] there is little chance of any change to the doctrine it espouses in the near future.

2 Direct liability

With the content provider's immunity under sec 230, defamation victims can only bring actions against the online speakers. This outcome might lead to a dead end, as such defendants, even when reachable, are not as deep-pocketed as some of the potential indirect defendants—large internet firms, such as Google, Yahoo or Facebook, which serve as content providers. Nonetheless, this is a legal path that victims of online defamation indeed (at times, reluctantly) pursue. In cases of anonymous speech, victims must first uncloak the primary wrongdoers' identities. To do so, they rely on the courts for assistance. Courts, in turn, might be in no rush to respond to this request. The right to anonymity is well recognised in US law. In some instances, especially when it pertains to speech and assembly, anonymity receives constitutional protection.[33] Thus, disclosing an anonymous user's identity is far from trivial. Nonetheless, a sufficiently stated cause of action can justify narrowly tailored and content-neutral regulation of anonymous speech, and exposure of anonymous speakers' identities.[34]

---

[27] However, empirical studies showed that more than one third of the relevant claims survived the sec 230 defence and accordingly websites had to engage in a long and expensive legal battle. See *DS Ardia*, Free Speech Savior or Shield for Scoundrels: An Empirical Study of Intermediary Immunity under Section 230 of the Communications Decency Act (2010) 43 Loyola of LA Law Review 373, 493; *Chander* (fn 8) 655.

[28] See *Chander* (fn 8) 653 n 58 for an extensive list of cases.

[29] *Id* at 651.

[30] See *DJ Solove*, The Future of Reputation: Gossip, Rumor, and Privacy on the Internet (2007) 159 (promoting a notice and takedown approach); see also *Ardia* (fn 27).

[31] *DK Citron*, Cyber Civil Rights (2008) 89 Boston University Law Review 61.

[32] See generally *Chander* (fn 8).

[33] *Froomkin* (fn 5) 442.

[34] *Id* at 447.

The key to uncloaking the relevant speaker is the 'John Doe Subpoena' issued by courts to relevant intermediaries.[35] Lawsuits against anonymous parties are a 'longstanding legal phenomenon,' often used to initiate a discovery process—seeking the defendant's true identity.[36] In the past, 'John Doe' processes were often used in civil rights lawsuits against unnamed police officers. As the internet age brought with it the ability to broadly engage in anonymous speech, this legal tool was quickly applied to the novel digital setting.

There is judicial controversy, and hence uncertainty, about the standard of evidence for establishing the plaintiff's claim,[37] which must be met prior to issuing such an order.[38] In *Cahill*,[39] the plaintiffs (Mr and Mrs Cahill) sought relief regarding comments posted on a local blog concerning Mr Cahill's performance as a City Councilman of Smyrna, noting his 'character flaws' and 'mental deterioration.' After seeking and obtaining the anonymous speaker's IP address from the blog owner, they approached Comcast to match the address to a specific user. The anonymous speaker filed a motion to prevent the disclosure of his identifying information, and thus the court was required to establish the required standard of evidence. The Delaware court held that a defamation plaintiff must support his claim with facts sufficient to defeat a 'summary judgment' motion before obtaining the identity of an anonymous defendant.'[40] In *Dendrite*,[41] a New Jersey court was called upon by a software company subjected to criticism on a Yahoo message board, and a similar standard of establishing a 'prima facie' case was applied.[42]

Yet other courts have applied the more lenient 'motion to dismiss' standard, whereby the claim must survive a motion to dismiss,[43] or the 'variable standard' which depends on the nature of the forum and the speech (commercial versus non-commercial).[44] In some states, specific

---

[35] *Gleicher* (fn 3) 325.

[36] *Id* at 327.

[37] See *Solove/Schwartz* (fn 17) 600; see also *Gleicher* (fn 3) 325, 337, 340 (identifying seven cases addressing distinct standards, adding the 'good faith standard,' a slightly altered 'summary judgment' rule in a case that involved trespass to chattels, and an altered 'prima facie' rule).

[38] A stricter standard was applied for 'unmasking' third parties to litigation. See *Doe v 2TheMart.com Inc*, 140 F Supp 2d 1088 (WD Wash 2001). For a discussion of the anonymity of witnesses in criminal cases, see *Froomkin* (fn 5) 449.

[39] *Doe No 1 v Cahill*, 884 Atlantic Reporter (A) 2d 451 (Del 2005).

[40] 'A judgment granted on a claim about which there is no genuine issue of material fact and upon which the movant is entitled to prevail as a matter of law.' Black's Law Dictionary (Abridged 7th edn 2000).

[41] *Dendrite International, Inc v Doe No 3*, 775 A.2d 756 (NJ Super 2001);

[42] 'A party's production of enough evidence to allow the fact-trier to infer the fact at issue and rule in the party's favor.' Black's Law Dictionary (Abridged 7th edn 2000).

[43] *Columbia Insurance Co v Seescandy.com*, 185 FRD 573 (ND Cal 1999) (a trademark infringement claim). This is a more lenient standard than that of 'summary judgment,' and merely requires pleading sufficient facts to survive a motion to dismiss, as opposed to providing prima facie evidence sufficient to withstand a motion for summary judgment.

[44] *Anonymous Online Speakers v US District Court of Nevada Reno*, 611 F 3d 653 (9th Cir 2010).

standards and processes for unmasking anonymous speakers have been set forth by the legislature. For instance, Virginia Code § 8.01-407.1 provides a detailed procedure for obtaining anonymous user information, and sets an even more lenient standard for plaintiffs (they only need to have 'good faith basis to contend' that they are victims of actionable conduct). Although this section was found constitutional in *Yelp, Inc v Hadeed Carpet Cleaning, Inc*, [45] the question of its constitutionality might eventually make its way to the US Supreme court.

Unsurprisingly, the breadth of the John Doe Subpoena has generated vibrant academic debate. Commentators noted that a standard which was over-protective of the anonymous speaker would enable excessive harassment.[46] On the other hand, a standard too lenient would compromise speakers' speech-related interests. In some instances, such as those involving whistleblowers, the effort to expose the speaker's identity (as opposed to the libel case that follows) is the most crucial and substantial legal battle. This is most evident where the plaintiff can take action against the defendant outside the courtroom, for example, by public shaming or by terminating their employer-employee relationship.

In some cases the anonymous speaker's identity is a key factor in establishing the defamatory nature of the statement. Therefore, a rule requiring the plaintiff to establish defamation prior to piercing the speaker's veil of anonymity places the former in an awkward position. For instance, in *Yelp v Hadeed*,[47] the plaintiff argued that anonymous customer reviews of its business were defamatory because the reviewers were not real customers voicing their genuine opinions but, rather, individuals falsely representing themselves as disappointed customers to harm its reputation. It based these allegations on the failure to match the critical reviewers—based on the information provided in their reviews—with real customers in its database. Yet determining if the speakers were real disgruntled clients required disclosure of their identities. The Virginia Court of Appeals, applying the lenient standard under the Virginia Code, found for the plaintiff, but the issue will soon be brought before the state's Supreme Court.

To summarise, even though the John Doe process is not without criticism, its existence is an established fact which allows plaintiffs to pursue the primary wrongdoer. In many instances, such action cannot be taken against the content provider in view of sec 230 of the CDA.

---

[45] *Yelp, Inc v Hadeed Carpet Cleaning, Inc,* 752 South Eastern Reporter (SE) 2d 554 (Va App 2014). For a discussion of other jurisdictions, see *id* at 562 n 6.

[46] *M Kaminski*, Real Masks and Real Name Policies: Applying Anti-Mask Case Law to Anonymous Online Speech (2013) 23 Fordham Intellectual Property, Media & Entertainment Law Journal 815, 828.

[47] Supra fn 45.

C Israeli model: exclusive indirect liability

1 Indirect liability

At present, Israel sets forth a unique liability model. Here, a claim against the anonymous alleged wrongdoer is *de facto* blocked, while a claim against the platform could proceed in some specific instances. The analysis of Israeli law is mostly premised on case law. Israel does not have a legal framework similar to that set forth by the EU Directive discussed below, although a bill that addressed the problem along the lines articulated in the Directive was introduced in 2008 and again in 2011.[48] At both times, however, the bill failed to pass.

Sec 11 of the Israeli Defamation Act provides that where the defamatory statement was published in a communications medium, the following parties may be held liable in addition to the person who made the statement: (1) the person who brought the defamatory matter to the newspaper and thus caused its publication; (2) the editor of the newspaper; (3) the person who made the decision to publish; and (4) the person responsible for the specific medium (for example, the publisher in the case of a newspaper). For purposes of the Defamation Act, 'communications media' include only newspapers (as defined in the Journalism Ordinance), radio, and television. In very limited cases courts have found content providers liable as 'communication media' because they resembled newspapers and their content was closely moderated.[49] . But it is settled that other websites are not 'newspapers' and sec 11 thus does not apply to publications on them.[50]

Beyond claims premised upon the Defamation Act, courts have contemplated the applicability of the general tort of negligence to content providers' conduct. At least on one occasion, the Israeli Supreme Court has noted that a dating website might be found liable for negligence. The Court reached this conclusion while referring to the website's poor handling of offensive anonymous comments, among others, given the lack of ex ante screening of content and identity verification for users; the Court even upheld an award of punitive damages.[51] Thus, Israeli law allows actions against content providers, yet only in well-defined cases.[52]

---

[48] Electronic Commerce Bill, 2008 Hazaot Hok (HH) 322, §§ 7-10 (concerning ISPs liability), 13 (concerning disclosure of personal information and uncloaking anonymous users); Electronic Commerce Bill (25 July 2011, Private Members' Bill), §§ 10 (regarding liability), 13 (regarding the disclosure of information and the uncloaking of identity).

[49] Civil Case 5844-07 (Magistrate Court Rishon LeZion) *Weintraub v Globes Publisher Journalism (1983) Ltd* (3 December 2009); see also *R Perry*, Israel, in: European Tort Law Yearbook 2011, 715, 732, no 41; *U Volovelsky/R Raynzilber*, The Liability of Website Owners for Defamation in Israel: A Challenge Yet to Be Solved? (2013) 29 Computer Law & Security Review 590, 599.

[50] See *Perry* (fn 49) 732, no 41; ; see also *Volovelsky/ Raynzilber* (fn 49) 596.

[51] Permission Civil Appeal 1700/10 *Dubitzky v Shapira* (Supreme Court, 20 May 2010); see discussion in *Perry* (fn 49) 731-733, nos 40-43.

[52] *Perry* (fn 49) 726, no 27.

2 Direct liability

Although the ability to bring actions against content providers is limited, this is currently the only course of action available to victims of online anonymous defamation in Israel. Dragging an anonymous online speaker to court is close to impossible. In the *Rami More* case, the Israeli Supreme Court found that there was no procedural framework for ordering intermediaries to provide identifying data about anonymous users.[53] In this ruling the court de facto shielded the specific anonymous speaker from tort liability. Israeli law thus takes a clear position about the uncloaking of anonymous speakers.[54]

At first sight the *Rami More* ruling may seem technical, expressing no substantive preference of the legal system. Yet this reading must be rejected. The Supreme Court's decision, effectively insulating the online anonymous speaker from liability, should be understood as one of substance for at least three reasons. First, the dissenting Justice, as well as lower courts—including the District court in this specific case, had no difficulty finding sources of authority to order disclosure of alleged tortfeasors' identities.[55] Yet the majority chose to reject these sources, as well as the suggestion to apply the English *Norwich Pharmacal* rule in Israel. [56] This relatively activist Court would not hide behind mere technicalities if it believed that a different result was warranted. Indeed, the Court backed its decision by noting the importance of maintaining anonymity as a reflection of both free speech and the right to privacy in the digital economy.[57]

Second, the legal system's response, as a whole, to the *Rami More* ruling also indicates its acceptance of the substantive outcome. The court explicitly noted that the legislature could intervene if online wrongdoers should be uncloaked.[58] However, even though four years have passed, and several bills have been set forth,[59] the legal reality has yet to be modified.

Third, the Supreme Court's position regarding the technical inability to bring legal action against online anonymous speakers was recently revisited and upheld in a related context—that of online anonymous copyright

---

[53] Permission Civil Appeal 4447/07 *Rami More v Barak ITC – International Telecommunications Corp*, 63(3) PD 664 (2010); see also *Perry* (fn 49) 724-725, no 24.

[54] Indeed, this exact point was recently articulated by the court in Civil Case 4854-07 (Magistrate Court Rishon LeZion) *Blumenfeld v Chobod* (22 April 2014), para 116.

[55] See Permission Civil Appeal 4447/07 *Rami More v Barak ITC – International Telecommunications Corp* (Supreme Court, 25 March 2010), Justice Rubinstein's dissenting opinion (particularly paras 4-5).

[56] *Id* para 34 (majority opinion).

[57] *Id* para 37 (majority opinion). This statement follows a longer discussion of the importance of anonymity in general and in the online realm in particular (paras 11-17).

[58] *Id* para 36 (majority opinion).

[59] See Disclosure of User Information in Electronic Communications Network Bill, 2011 HH 36; Disclosure of User Information in Electronic Communications Network Bill, 2012 HH 1376. See discussion in *Perry* (fn 49) 724-726, nos 24-27; *R Perry*, Israel, in: European Tort Law Yearbook 2012, 745, 751-752, nos 16-18.

infringement.[60] Here, the majority of the court again found that given the lack of a relevant legal apparatus, it could not grant an order to reveal the identity of the anonymous speaker. The Justices indicated that they might be inclined to creatively use existing legal doctrines to enable uncloaking of alleged copyright infringers should the legislative response fail to materialise in the near future. However, these statements explicitly pertained to copyright infringement. It is fair to assume that the court will *not* take an equally aggressive stance in the context of online anonymous defamation, given the speech-related interests involved.[61] One of the court's senior Justices opined that the law must afford greater protection to the right of anonymous speech in the context of defamation than in the context of copyright infringement.[62] The explicit distinction between copyright and defamation cases constitutes additional support for the view that the court's ruling in *More* reflects a substantive preference, rather than a mere procedural lacuna. Thus, the court should be understood to take a normative position regarding the ability to take action against the direct wrongdoer in these cases.

D EU framework: concurrent liability

1 Indirect liability

At this time, there are no relevant EU Regulations. Therefore, a comprehensive analysis of the law applicable to the issues at hand calls for a separate examination of national law in each Member State. State laws point in different directions and apply different strategies to address the subject.[63] Such a project is beyond the limited scope of this article. Instead, we strive to grasp the contours of the European framework. These are drawn by the E-Commerce Directive, and by a very recent decision of the European Court of Human Rights (ECtHR) in the case of *Delfi AS v Estonia*.[64] The two sources define a general model which can be compared to the alternatives even without delving into the intricacies of its implementation in each Member State.

The somewhat dated E-Commerce Directive (the 'Directive')[65] sets up a framework for electronic commerce, and establishes harmonised rules to promote legal certainty for businesses and consumers.[66] In that way it aims

---

[60] Civ App 9183/09 *The Football Association Premier League v John Doe* (Supreme Court, 13 May 2012).

[61] *Perry*, ETL 2012 (fn 59) 771, no 65.

[62] *Id* at 769-770, no 62.

[63] *Verbiest/Spindler/Riccio/Van der Perre* (fn 22) 14 ('National implementation and court practice differ between Member States considerably when assessing actual knowledge').

[64] *Delfi AS v Estonia* [2013] ECtHR, Application No 64569/09 (10 October 2013).

[65] Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market ('Directive on electronic commerce').

[66] Introduction to the Directive <http://ec.europa.eu/internal_market/e-commerce/directive/index_en.htm>.

to support the growth of the online industry. Inter alia, the Directive draws out broad rules which pertain to a variety of speech related torts, such as defamation and IP infringement (the latter is not within the ambits of this article). The Directive merely sets a general threshold that all states must comply with.

The overall strategy is very different from that applied in the US. As Anupam Chander recently noted, the Directive 'stops far short of the near blanket exclusion from liability offered by the Communications Decency Act.'[67] Section 4 of chap II addresses the issue of 'Liability of intermediary service providers.' It starts out, in art 12, by providing immunity to intermediaries which act as mere conduits of information, which are not discussed in this article. In addition, art 14 provides that intermediaries engaged in 'hosting' should not be liable, unless they have actual knowledge of the illegal statements, or refuse to expeditiously remove them upon obtaining knowledge. In other words, hosts are subject to a 'notice-and-takedown' rule, namely an obligation to remove problematic content once a potential plaintiff brings its existence to the intermediary's attention.[68] According to art 15(1), Member States should not impose a general obligation on mere conduits and hosts to monitor the content they transmit or store.

These provisions can result in liability should the court find that a content provider which may be regarded as a 'host' failed to promptly comply with takedown requests or had actual knowledge of the wrongdoing. Moreover, the Directive Recitals (47 and 48), indicate that Member States can impose specific—as opposed to general—monitoring obligations and, even more importantly, subject hosts to duties of care, 'which can reasonably be expected from them and which are specified by national law, in order to detect and prevent certain types of illegal activities.'[69] Lastly, the Directive is somewhat dated and its classification might no longer be comprehensive. Many content providers may not be 'hosts' at all, as we will shortly demonstrate. In such cases, the Directive's restrictions do not apply. In summary, the EU framework enables a defamation victim, in some instances, to bring action against the content provider.

Beyond the E-Commerce Directive, the potential liability of an online platform was recently addressed by the ECtHR, in the case of *Delfi AS v Estonia*.[70] Here, the court upheld the Estonian Court's ruling, finding the popular Delfi news website liable for defamatory statements about a famous Estonian business executive, posted by anonymous users on that website in response to an article discussing his recent business ventures.

---

[67] *Chander* (fn 8) 671.

[68] E-Commerce Directive, Art 15. For the variance in application among Member States, see *Verbiest/Spindler/Riccio/Van der Perre* (fn 22) 16; see also infra notes 153-164 and accompanying text.

[69] E-Commerce Directive, Recitals 47-48; *Verbiest/Spindler/Riccio/Van der Perre* (fn 22) 5.

[70] *Delfi AS v Estonia* [2013] ECtHR, Application No 64569/09 (10 October 2013).

The court found Delfi liable although the website used a system for automatically deleting specific terms and clearly articulated in its 'terms of use' that profanity was forbidden, and that any content posted did not reflect its opinion. The court acknowledged that Delfi's comment section was notorious for its defaming content, a fact which might have contributed to the court's final decision. Most importantly, Delfi was found liable even though it applied a 'notice and takedown' process and thus complied with the abovementioned requirements of the EU Directive. The ECtHR agreed with the Estonian Supreme Court that Delfi could not benefit from arts 12-14 of the Directive (and Estonian law which implemented them), because both Delfi and the authors of the defamatory comments were to be considered publishers of these comments.[71] Delfi's actions rendered it a 'publisher,' rather than a mere 'intermediary.' Therefore, it was not relieved from 'monitoring' online posts and could not be exempted from liability by taking down content. It should come as no surprise that the *Delfi* decision generated substantial confusion as to the distinction between online 'publishers' and mere 'intermediaries' and the extent of legal protection that adherence to a 'notice and takedown' process provides.[72]

While this ruling is based on the interpretation of a relevant Estonian statute, it illuminates the stark contrast between content providers' liability in the US and in the EU. While, as noted above, US law emphasises the primacy of promoting online discourse and freedom, the ECtHR's opinion emphasised that free speech, on the one hand, and the protection of honour and personhood (which may be compromised in the absence of liability), on the other hand, should be equally respected by the court.[73] This different balance of human rights no doubt led to the final outcome, which recognises extensive liability of content providers.

2 Direct liability

The EU framework thus enables legal actions against content providers. It also enables actions against anonymous speakers who are the primary wrongdoers. To examine the feasibility of uncloaking anonymous users we return to the Directive. Article 15(2) allows Member States to establish obligations for service providers to transfer users' identifying information to competent authorities, including courts. Disclosure

---

[71] *Id* para 28; see also para 50 ('The Court notes that the applicant company was sued for defamation in respect of comments posted on its Internet portal, it was deemed to be discloser (or publisher – the Estonian words avaldama/avaldaja mean both disclose/discloser and publish/publisher; see, for example, paragraphs 36 and 38 above) of the comments – along with their authors – and held liable for its failure to prevent the disclosure of or remove on its own initiative the unlawful comment.').

[72] See *G Smith*, Who Will Sort Out the Delfi Mess (16 October 2013) Cyberleagle <http://cyberleagle.blogspot.com/2013/10/who-will-sort-out-delfi-mess.html>.

[73] *Delfi AS v Estonia* [2013] ECtHR, Application No 64569/09 (10 October 2013), para 82.

processes should comply with the Data Protection Directive,[74] the Electronic Privacy Directive,[75] and national data protection laws, rendering them complex and state specific.[76]

A report studying the implementation of the Directive[77] draws out a complicated mapping of state rules which pertain to requests to provide identifying information addressed to content providers and other internet intermediaries.[78] The report reviews a variety of instances in which plaintiffs requested orders that would provide them with identifying information about defendants, and such orders were granted. The majority of these cases pertain to intellectual property enforcement. This, most likely, results from the fact that IP right holders have both the resources and the strong interest to take immediate action against possible infringers. However, there are also examples for disclosure orders in the defamation context. For instance, at the request of Ryanair, the Irish High Court issued an order requiring Eircom, the Irish national telecommunications provider, to disclose the identities of anonymous users who posted defamatory comments regarding the Irish-based airline (Ryanair).[79]

To summarise, the EU approach allows plaintiffs to take legal action along two parallel routes in cases of online anonymous defamation; they are provided with legal measures to pursue both an anonymous speaker and the content provider. This overall outcome is, most likely, not accidental but rooted in an ideological commitment. Such ideology is reflected the ECtHR's *Delfi* decision, where the court specifically noted that the law could simultaneously allow actions against the speaker and the platform.[80] It explained that limiting plaintiffs' reach to merely suing speakers (the position taken in the US) would not provide potential victims with effective

---

[74] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, Official Journal L 281, 23/11/1995 P 0031-0050.

[75] Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector ('Directive on privacy and electronic communications'), Official Journal L 201, 31/07/2002 P 0037-0047.

[76] *Verbiest/Spindler/Riccio/Van der Perre* (fn 22) 80.

[77] *Id* at 75-79.

[78] For instance, the report notes the reluctance of German courts to allow intermediaries to provide such information (*id* at 76-77). However, since the publication of the report, German courts have formulated ways to order the uncloaking of online speakers, although some of the issues related to data protection are yet to be resolved (private communication with Gerald Spindler, 11 April 2014, on file with authors).

[79] *A Loughlin*, Ryanair Seeks to ID Defamatory Online Parties (13 February 2013) Irish Examiner <http://www.irishexaminer.com/ireland/ryanair-seeks-to-id-defamatory-online-parties-222493.html>. For further discussion of Irish case law see *Verbiest/Spindler/Riccio/Van der Perre* (fn 22) 78-79. Note that Irish law makes use of *Norwich Pharmacal* orders to uncloak anonymous speakers—a measure originating in the UK (and discussed at length in the next section). It therefore might not prove an ideal example of the EU framework, as other EU Member States do not feature this common law mechanism.

[80] *Delfi AS v Estonia* [2013] ECtHR, Application No 64569/09 (10 October 2013), para 77.

protection.[81] Nevertheless, we take note that in the specific case discussed the court allowed an action against the intermediary where establishing the identities of the anonymous authors was very difficult.[82]

E English model: direct liability and residual indirect liability

1 Indirect liability

England presents a unique model in view of the newly minted Defamation Act of 2013. This law was given Royal Assent on April 25, 2013, and came into force on January 1, 2014. It joins a complex set of statutory and common law. Given this Act's novelty, it is quite difficult to make conclusive statements about its interpretation and application. Yet on its face, it provides an interesting combination of the two components addressed in this article: the content provider's liability and the ability to bring legal action against the online anonymous speaker. Rather than addressing them as independent elements, they are interconnected.

Prior to delving into the 2013 Act, we must note that pre-existing English law allowed one to bring actions both against the content provider and against the anonymous speaker. The law pertaining to the latter is most likely unchanged by the new Act, and will be discussed below. With regard to the former, the relevant background is perhaps most clearly manifested in the recent case of *Tamiz v Google*,[83] which was decided shortly before the enactment of the new Act. Tamiz was a conservative politician running for office in local elections. Blog posts and comments accused him of drug dealing, theft from his employer, and hypocrisy. The Court of Appeal held that the blog platform operator (in this case, Google) should be considered a 'publisher' and may therefore be found liable for defamatory blog posts after receiving an abuse report. Ultimately, the court found that given the limited time between notification and removal there had been no 'real and substantial tort.'[84] However, the broad recognition of publisher liability in this setting provided an important lesson for other potential plaintiffs.

Arguably, the extent of content providers' liability was somewhat altered by the 2013 Act. The premise that content providers can be found liable is most likely unchanged. But various actions taken by defamation-victims and content providers might impact the outcome. Among the Defamation Act's innovative legal mechanisms, those drawn out in secs 5 and 10 are relevant to our discussion of content providers' liability. The practical role these provisions will play in governing online intermediaries, as well as the problematic interaction between them, are still unclear.[85] Yet

---

[81] *Id* para 91.

[82] *Id*.

[83] *Tamiz v Google* [2013] EWCA Civ 68.

[84] See *D Rolph*, Defamation by Social Media (2013) 117 Precedent 16.

[85] *A Mulis/A Scott*, Tilting at Windmills: the Defamation Act 2013 (2014), 77 Modern Law Review 87, 101; *I Wilson*, The Defamation Act 2013 (24 February 2014) The Law Society Gazette <http://www.lawgazette.co.uk/law/legal-updates/the-defamation-act-2013/5039959.article> ('it is too early to assess the impact of sections 5 and 10.').

as we shall now explain, both can be considered as promoting the idea that a content provider should be liable only when the primary wrongdoer is unreachable.

Section 5 explicitly applies to websites, as its title—'operators of websites'—evinces. A careful reading of this section can lead to the conclusion that it endorses the idea that a victim who can bring an action against the speaker should not be permitted to sue the content provider as well. Section 5 provides that a website operator is generally not liable for a defamatory statement posted on the website if it was not the one who posted that statement.[86] The defence can be defeated, however, if the victim has insufficient information to identify and bring proceedings against the speaker,[87] the victim gave notice of complaint, and the content provider did not respond to the victim's complaint in accordance with the Defamation (Operators of Websites) Regulations. If the content provider cannot contact the speaker, if the speaker does not respond or does not provide the required information (including personal name and address), or if he or she agrees to removal, the allegedly defamatory content should be removed immediately. Otherwise, the content provider need not remove the content, and may provide the speaker's contact information to the victim if the former consents or if a court orders it to do so. The process articulated in sec 5 is voluntary. The intermediary could ignore it and take its chances in court. It might also rely on other defences recognised in English law.[88]

Under sec 5, therefore, the content provider's liability is residual: it depends on non-removal where the speaker is unreachable. There is no indirect liability if the speaker's identity is known or if the content provider takes steps to make an action against the speaker possible. By rejecting the idea that the law should enable concurrent actions against the speaker and against the content provider, the new Act challenges the policy advocated by the ECtHR in *Delfi*.

Section 10 of the new Defamation Act could also be of relevance. This section, which does not specifically apply to online platforms, addresses the relation between direct and indirect liability for defamation, and is therefore related to our discussion. Section 10 provides that a 'court does not have jurisdiction to hear and determine an action for defamation brought against a person who was not the author, editor or publisher of the statement complained of unless the court is satisfied that it is not reasonably practicable for an action to be brought against the author, editor or publisher.'[89] Arguably, a victim cannot sue the content provider if the

[86] Sec 5(2).

[87] Secs 5(3)(a), 5(4).

[88] Eg, Regulation 19 of the Electronic Commerce (EC Directive Regulations) 2002, or the defences provided in sec 1 of the Defamation Act 1996. See *J Agate*, The Defamation Act 2013 – Key Changes for Online (Farrer & Co, September 2013) <www.farrer.co.uk/Global/Briefings/-06 Private Client/The Defamation Act 2013 - key changes for online.pdf>.

[89] These terms carry the same meaning as in the Defamation Act, 1996 c 31.

speaker can be reached through reasonable effort.[90] So content providers' liability may be residual.[91]

This last assertion comes with two caveats. First, sec 10 only pertains to instances in which the content provider cannot be considered an 'editor' or a 'publisher.' If and when online content providers are considered editors or publishers (as opposed to distributors), however, sec 10 does not apply. As noted above, in *Tamiz*, the court took an expansive view of the term 'publisher,' potentially narrowing the applicability of sec 10 to online content providers. However, several commentators explain that online content providers should rarely be recognised as editors or publishers.[92] Such recognition might only be established once the relevant websites are put on proper notice that potentially defamatory content was posted, as in *Tamiz*. Section 10 applies in all other instances.

Second, the meaning of a key term in sec 10, 'reasonably practicable,' remains unclear.[93] In the context of our discussion, courts must still clarify whether the legal and technical steps required to bring said speaker to court are prima facie reasonably practical. Should courts find that carrying out the legal and technical steps to uncloak the online anonymous speaker present an unreasonable burden, sec 10 will prove irrelevant. Such a finding will in fact free plaintiffs to pursue legal actions against both speakers and content providers, in a way similar to the one noted in our discussion of general EU law (subject to sec 5, as discussed above). However, while identifying anonymous speakers is a tedious task, existing precedents make it potentially 'reasonable.' Indeed, English law provides a tool for requiring internet intermediaries to help uncloak anonymous speakers, namely, the Norwich Pharmacal order. So it is possible that sec 10 precludes content providers' liability in the absence of a real obstacle to an action against the anonymous speaker.

2 Direct liability

Let us now address the second component in our analysis, namely the ability to pursue an action against the speaker. English law offers a well-established process, which was endorsed throughout the British Commonwealth,[94] for uncloaking anonymous wrongdoers—the Norwich

---

[90] Cf *Mulis/Scott* (fn 85) 101 (noting that this position is arguable, but it might 'render the section 5 notice procedure redundant, and that is presumably not what was intended by Parliament').

[91] *Id* (noting that sec 10 makes it significantly more difficult to proceed against online intermediaries).

[92] *N Armstrong*, Briefing Note: Defamation Act 2013, Update: Website Operators (Charles Russell LLP, September 2013) 4 <http://www.charlesrussell.co.uk/UserFiles/file/pdf/Reputation Management/Defamation_Act_Website.pdf>/

[93] Eg, do they also include instances in which the direct tortfeasor is insolvent? *Mulis/Scott* (fn 85) 101.

[94] See Permission Civil Appeal 4447/07 *Rami More v Barak ITC – International Telecommunications Corp* (Supreme Court, 25 March 2010), para 34.

Pharmacal order. This order is named after a 1974 case[95] which established that an innocent party mixed up in another's tortuous acts has a duty to provide information to the victim.[96] The context in which this legal tool was first applied was quite different from the one discussed here: a patent holder requested that the customs authorities provide information about an alleged infringing importer.[97]

Yet in recent years, this legal measure was applied to different types of internet intermediaries,[98] in a variety of contexts, including anonymous defamation.[99] The case law has developed various criteria and prerequisites which must be met prior to granting the order. However, it is clear that existing law provides plaintiffs with the ability to drag anonymous online speakers into the courtroom, should they choose to do so. Section 5 of the 2013 Act and the Regulations might even simplify this process.

Illustrative, in a recent decision that received significant media attention[100] the High Court of Justice in Northern Ireland ordered that unmasked formerly-anonymous speakers pay a substantial sum of damages for posting false messages about the plaintiffs on Facebook. The court somewhat proudly noted that even though the speakers attempted to cloak their identities, they were nonetheless brought to court to answer for their actions, thanks to flexible and appropriate legal measures. In the court's words:

> [O]ur legal system, mainly through the media of increasingly flexible procedural rules and practices, resorts to the inherent jurisdiction of the High Court where appropriate and a progressive emphasis on substance in preference to form adopts a robust and realistic approach to issues of this kind. As a result, the defendants' attempts at concealment and evasion have been thwarted. I am satisfied that they are properly before and, hence, are subject to the jurisdiction of this court.[101]

III. ECONOMIC ANALYSIS

A Overview

This Part analyses the competing legal regimes from an economic perspective. In our model there are two potential defendants in an action for

---

[95] *Norwich Pharmacal Co v Commissioners of Customs & Excise*, [1974] AC 133 (HL).

[96] *BG Baynham/DJ Reid*, The Modern Day Soapbox: Defamation in the Age of the Internet (Continuing Legal Education Society of British Columbia, September 2010) 3.1.11 <https://www.cle.bc.ca/PracticePoints/LIT/11-ModernSoapbox.pdf>

[97] See *M Birnhack*, Private Space (2011) 358 (in Hebrew).

[98] It is available vis-à-vis content providers, as well as mere conduits. For an example of the former, see *Lockton Companies International & Others v Persons Unknown and Google Inc*, [2009] EWHC 3423 (QB). For a discussion on the latter, see *DM Síthigh*, The Fragmentation of Intermediary Liability in the UK (2013) 8 Journal of Intellectual Property Law & Practice 521, 522.

[99] For an early example, see *Totalise PLC v The Motley Fool Ltd*, [2001] EWHC 706 (QB).

[100] *AB Ltd v Facebook Ireland Ltd*, [2013] NIQB 14 (6 February 2013).

[101] *Id* para 3.

online anonymous defamation: the speaker and the content provider. In theory there may be additional defendants, such as internet access providers.[102] But we focus here solely on the strongest candidates for liability. The speaker and the content provider may be liable or non-liable, either because substantive law does not impose liability (a substantive impediment) or because a liability rule cannot be enforced (a procedural impediment). Thus, there are four possible legal regimes: (1) no liability at all; (2) exclusive direct liability, namely liability of the speaker, the primary wrongdoer; (3) exclusive indirect liability, namely liability of the content provider; and (4) some combination of direct and indirect liability. A combination of direct and indirect liability can take several forms. For example, the speaker and the content provider may be jointly and severally liable in any case of online defamation. Alternatively, the content provider may be liable only if the speaker cannot be identified, so that indirect liability is residual. We assume that 'no liability at all' (the first legal regime) is unwarranted, because it eliminates all incentives to take precautions in creating, publishing, and disseminating potentially defamatory statements. We shall therefore discuss the remaining options.

B Exclusive direct liability

1 The basic justification

A speaker's liability for defamation is a special case of direct tort liability. Its justifications are, therefore, identical to the traditional justifications for direct liability in tort. Admittedly, scholars have set forth many non-economic justifications, including corrective justice,[103] civil recourse,[104] and possibly even distributive justice.[105] But we focus here on the economic perspective, with several exceptions. By holding the actual doer liable, the law can force potential doers to internalise the external costs of their activities, thereby inducing them to choose the socially desirable levels of care and activity.[106] Ideally, imposing liability on the primary wrongdoer should secure efficient conduct. However, in cases of online anonymous defamation, this simple model might not work. In what follows we discuss the reasons.

---

[102] These play a role in the process of identifying anonymous speakers.

[103] *JL Coleman*, The Practice of Corrective Justice (1995) 37 Arizona Law Review 15.

[104] *BC Zipursky*, Rights, Wrongs, and Recourse in the Law of Torts (1998) 51 Vanderbilt Law Review 1.

[105] *GC Keating*, Distributive and Corrective Justice in the Tort Law of Accidents (2000) 74 Southern California Law Review 193.

[106] *JCP Goldberg*, Twentieth-Century Tort Theory (2003) 91 Georgetown Law Journal 513, 544-553.

2 Identifying and pursuing the wrongdoer

The first problem is that an action for online anonymous defamation entails special effort in identifying and pursuing the wrongdoer.[107] At times, identifying the wrongdoer would be practically impossible. Even when possible, the cost of identifying an anonymous wrongdoer might be prohibitively high, dissuading the victim from bringing an action. The prospect of no-action reduces potential wrongdoers' expected liability, prevents full internalisation, and impairs their incentive to act efficiently. Finally, the victim may have an incentive to sue if the expected benefit offsets his or her personal administrative costs, even if they are substantial. However, in deciding whether to sue the victim takes into account only the expected costs and benefits for himself or herself, and ignores the costs incurred by other parties involved in the litigation. An action might prove inefficient if the aggregate administrative cost, including that incurred by internet service providers and the courts, outweighs the social benefit in terms of deterrence.

The costs of identifying an anonymous cyber-wrongdoer are prohibitively high due to legal and technological constraints. To identify the anonymous speaker the victim first needs the content provider to divulge the speaker's internet protocol (IP) address. Then, the victim needs the anonymous speaker's internet access provider, as identified by the IP address, to disclose the identity of the person using this address when the defamatory statement was made. These two steps raise serious constitutional and legal concerns, because of the potential infringement of the freedom of speech and the right to privacy, and therefore cannot be carried out without explicit authorisation under the law. Without a legal obligation to disclose users' personal information, an action against an anonymous cyber-wrongdoer is impossible, as the Israeli experience clearly demonstrates. A formal procedural framework for exposing the identity of anonymous cyber-wrongdoers, which takes into account the relevant constitutional and legal concerns, would be complex and entail high administrative costs.

Moreover, even when the law mandates disclosure, several obstacles arise. First, an action against the speaker may be impossible if the information disclosed by the content provider does not enable immediate identification. Sophisticated users who wish to preserve anonymity can easily hide their IP addresses, for example, by using anonymising proxy servers or Tor.[108] In such cases, the content provider is unable to disclose the speaker's true IP address. Thus, while some governmental agencies can track down these users, the average tort plaintiff cannot.

---

[107] *Hamdani* (fn 6) 910; *LB Lidsky*, Silencing John Doe: Defamation & Discourse in Cyberspace (2000) 49 Duke Law Journal 855, 869-870.

[108] *D Lichtman/E Posner*, Holding Internet Service Providers Accountable (2006) 14 Supreme Court Economic Review 221, 234 (explaining that sophisticated wrongdoers can conceal their tracks 'by routing messages through a convoluted path that is difficult for authorities to uncover').

Second, an IP address cannot always be attributed to a single user. For instance, if thirty people connect to a wireless network at a coffee shop, a library or any other public location, all will normally share the same IP address. In that case, the content provider can disclose a genuine IP address—that of the coffee shop—but the coffee shop's internet access provider cannot identify the specific wrongdoer. Tracking anonymous wrongdoers may be extremely expensive. For example, if a defamatory statement was made by a coffee shop user, a thorough investigation may lead to the actual speaker, but at a very high cost. The problem may be alleviated by imposing liability on hot spot operators, as is the case in Germany,[109] thereby incentivising them to require user identification.

Third, an action against the speaker may be impossible if the content provider or the speaker's internet access provider do not keep the users' log for a long enough period (as in the *Zeran* case). The law can once again intervene by requiring logs to be kept for a sufficiently long time. This type of regulation is commonly known as 'data retention laws.' Retention of information has a cost that is correlated with the daily traffic and the required duration of retention. More importantly, retention laws should not infringe basic rights. On 8 April 2014, the European Court of Justice held that the EU Data Retention Directive,[110] which required telecom companies to store user data for up to two years, was invalid.[111] It opined that '[b]y requiring the retention of the data… and by allowing the competent national authorities to access those data, Directive 2006/24' infringes the right to privacy and the right to the protection of personal data.[112] By adopting Directive 2006/24, the EU legislature has exceeded the limits imposed by compliance with the principle of proportionality.[113]

Fourth, a legal disclosure mechanism would often have territorial application, enabling anonymous speakers who make defamatory statements on foreign websites or through foreign ISPs to get off scot-free. But even where international arrangements enable extra-territorial disclosure orders, the process vis-à-vis foreign entities may be wearisome and very costly. Although we assumed for simplicity that all parties are subject to the same jurisdiction, this possible complication is too important to ignore.

---

[109] *M Rosenbach/H Schmundt*, Radio Silence: Germany's Wireless Internet Problem (4 July 2013) Spiegel Online International <http://www.spiegel.de/international/world/free-wifi-shortage-german-laws-make-it-hard-to-provide-wireless-internet-a-909288.html> (explaining that 'German law holds the operator of a public hotspot liable for everything its users do online').

[110] Directive 2006/24/EC of the European Parliament and of the Council of 15 March 2006 on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or public communications networks and amending Directive 2002/58/EC (OJ 2006 L105, p. 54).

[111] *Digital Rights Ireland Ltd v Minister for Communications* (ECJ, 8 April 2014) <http://curia.europa.eu/juris/documents.jsf?num=C-293/12> (following a request by the Austrian Constitutional Court and the Irish High Court).

[112] *Id* paras 32-37.

[113] *Id* para 69.

The internet-facilitated possibility of republishing and reposting defamatory statements may add another layer of complexity and additional costs. Suppose that one person posts a defamatory text and others repost it without reasonable screening. If we can attribute to specific users ascertainable expansions of the statement's reach, for example, to new and distinct groups, each must be sued for the marginal harm he or she caused. Yet pursuing each wrongdoer entails special costs, making enforcement not-worthwhile for defamation victims and resulting in under-deterrence.[114] Although the problem of multiple-users causing independent harms might not be as acute as in the case of copyright violations, any impediment to enforcement impairs deterrence.

In summary, identifying the online anonymous speaker might be very costly. If the speaker is not identified, the costs of defamation are not fully internalised, and potential wrongdoers are not efficiently deterred. If, on the other hand, the speaker is identified through a costly process, wrongdoers internalise the costs of their wrongdoing, but the administrative cost may outweigh the benefit in terms of cost-reducing deterrence. Alternatively, the high administrative costs associated with identifying the primary wrongdoer might render another party (for example, the content provider) a more cost-effective target for enforcement efforts.

The standard economic solution for under-enforcement problems is to expand the wrongdoer's liability by multiplying the actual level of losses by the reciprocal of the probability of enforcement under the circumstances.[115] For example, if the probability of successful enforcement is 1/5, the extent of the injurer's liability should be determined by multiplying the actual loss by 5. Quite naturally, this method was advocated by economists as a proper solution for the identification problem in cases of online anonymous speech.[116]

Still, this method presents at least four difficulties. First, in many jurisdictions there is no doctrinal basis for awarding extra-compensatory damages; even where punitive damages are recognised, they are reserved for outrageous conduct,[117] and constitutional constraints, particularly those related to free speech, may prevent the economically desirable expansion of

---

[114] Cf *D Lichtman/W Landes*, Indirect Liability for Copyright Infringement: An Economic Perspective (2003) 16 Harvard Journal of Law and Technology 395, 397 (discussing this problem with respect to online copyright infringement).

[115] See *WM Landes/RA Posner*, The Economic Structure of Tort Law (1987) 16; *S Shavell*, Economic Analysis of Accident Law (1987) 148; *RD Cooter*, Punitive Damages for Deterrence: When and How Much? (1989) 40 Alabama Law Review 1143, 1148; *AM Polinsky/S Shavell*, Punitive Damages: An Economic Analysis (1998) 111 Harvard Law Review 870, 873–874; *CM Sharkey*, Punitive Damages as Societal Damages (2003) 113 Yale Law Journal 347, 365–368.

[116] *KJ Arrow et al*, Amicus Brief, Metro-Goldwyn-Mayer Studios Inc v Grokster Ltd (US Supreme Court) 5 <http://www.copyright.gov/docs/mgm/kenneth-Arrow-et-al.pdf>.

[117] See, eg, Restatement (Second) of Torts § 908 (1977) ('Punitive damages may be awarded for conduct that is outrageous, because of the defendant's evil motive or his reckless indifference to the rights of others'); *Zipursky* (fn 104) 95–96.

the damages.[118] For instance, if only one of every hundred anonymous speakers can be identified, it seems improper to multiply the harm by a hundred. Second, the proposed solution compels the court to assess the probability of escaping liability, and this may be impractical.[119] An incorrect assessment of the probability of escaping liability may result in over-deterrence or under-deterrence, depending on the nature of the error.[120] Third, even where it is possible to assess the probability of escaping liability, awarding extra-compensatory damages in accordance with the economic formula may aggravate the problem of judgment-proof defendants to be discussed below. Fourth, if wrongdoers in a particular category consistently evade liability, the multiplier method cannot work. We believe that in the context of online anonymous defamation, sophisticated—and particularly intentional—wrongdoers can consistently evade liability.[121]

3 Judgment-proof wrongdoers

An additional problem marring the effectiveness of the incentives provided by direct liability is the high likelihood of judgment-proof defendants. If the potential injurer may be unable to fully compensate for harm that may be caused by her conduct, she will not internalise the social cost of that conduct. From her perspective, the expected expense may be considerably lower than the expected (social) harm. So the incentive for choosing the optimal level of care is impaired.[122] For example, assume that there is a probability of 0.2 that A's conduct will cause a €10,000 loss to B, and that A can reduce the probability of harm to 0.1 by adopting a certain precaution for €800 (for example, verifying the accuracy of a factual statement). The cost of care (€800) is lower than the ensuing reduction in expected harm (€1,000), so from an economic perspective taking this precaution is desirable. Now assume that the expected value of A's assets during the subsequent litigation is €3,000, and that A is risk-neutral. Even if liability is certain, it will not provide an adequate incentive for choosing the optimal level of care. The expected sanction that would be imposed on

[118] See, eg, *State Farm Mutual Automobile Insurance Co v Campbell*, 538 US 408, 425 (2003) (holding that a single-digit ratio between non-compensatory and compensatory damages is more likely to accord with Due Process than awards with ratios in the range of 500 to 1 or even 145 to 1). Freedom of speech may call for an even more restrictive approach.

[119] *Shavell* (fn 115) 148.

[120] Indeed, the probability of under-enforcement (due to detection problems and unwillingness to sue) may be a source of controversy even in relatively simple cases. See *S Shavell*, On the Proper Magnitude of Punitive Damages: *Mathias v Accor Economy Lodging, Inc* (2007) 120 Harvard Law Review 1223, 1223–1227 (discussing Judge Posner's analysis of the magnitude of punitive damages in *Mathias*).

[121] Perhaps the law must handle this group separately.

[122] See *KD Logue*, Solving the Judgment-Proof Problem (1994) 72 Texas Law Review 1375, 1375; *S Shavell*, The Judgment Proof Problem (1986) 6 International Review of Law & Economics 45, 45; Comment, The Case of the Disappearing Defendant: An Economic Analysis (1983) 132 University of Pennsylvania Law Review 145, 157–159.

A for failing to take the optimal level of care would be only 0.2×€3,000=€600, whereas the cost of the precaution is €800 (and under a strict liability rule, this adds to expected liability of 0.1×€3,000=€300). Simply put, the wrongdoer's ability to pay is a de facto cap on liability and internalisation.[123]

Although the judgment-proof defendant is a general problem in tort law, it is particularly common in cases of online speech torts. Almost everyone in the developed world uses the Internet. The ease of access and the veil of anonymity encourage everyone to participate. So the typical user is essentially the average citizen with average assets and average income. Consequently, online speakers are often judgment-proof individuals.[124] The typical internet user may not have sufficient assets to pay for harm caused by his or her defamatory statement.[125] Thus, even if identification costs were low, exclusive direct liability for online defamation would result in under-deterrence. Obviously, a wrongdoer's inability to fully pay for harm caused results not only in under-deterrence but also in under-compensation. Exclusive direct liability may thus leave the victim with partly-compensated harm.[126] Although under-compensation is distinct from under-deterrence, and seems to be mostly relevant from a corrective justice perspective, it may have economic repercussions to the extent that full compensation is required to prevent expansion of initial harm.

The proponents of the economic approach to tort law traditionally propose several solutions to this problem. The first and most relevant in the current context is indirect liability: if there is a third party with sufficient financial resources who has some control over the potential wrongdoer's conduct, that party may be held vicariously liable for the conduct of the primary wrongdoer,[127] or liable for negligently failing to prevent it. In many cases there is no such person or organisation, and, even where such a person exists, the degree of control is rarely sufficient to ensure the efficient conduct of judgment-proof actors. Yet in the case of online defamation the content provider seems to be a natural candidate for indirect liability. Content providers have some control over user-generated content, and frequently have the resources to compensate defamation victims.

Another possible solution is to prohibit people who do not have the financial ability to cover losses that may be caused by certain activities from engaging in those activities. But this would be inefficient. From an economic perspective, a person should be allowed to perform actions the benefits of which exceed their costs. Setting a minimum asset requirement may curtail economically desirable activities. For example, an activity that yields a benefit of €11,000 and creates a 1% risk of a million-dollar loss, is efficient. But very few people will engage in it if they must have a million

---

[123] *Lichtman/Posner* (fn 108) 230.
[124] *Hamdani* (fn 6) 910-911.
[125] *Lichtman/Posner* (fn 108) 234.
[126] *Lidsky* (fn 107) 870.
[127] *Shavell* (fn 115) 168–169.

dollars to do so.[128] In the online speech context, if people would need to have sufficient funds for compensating defamation victims in order to use the internet, web access—and the ability to exercise free speech—would be limited to affluent individuals and corporations. The weak will be silenced, and socio-economic inequalities will be preserved and probably amplified.

A third solution is to require those engaged in particular activities to purchase liability insurance.[129] In fact, according to the EFF website, many insurance companies offer liability insurance policies designed to cover online defamation claims.[130] Arguably, homeowner's insurance policies, and possibly rental or umbrella insurance policies, may cover liability for defamation.[131] However, requiring insurance may once again curtail a free exchange of diverse ideas by excluding those unable to purchase insurance. Moreover, while insurance may guarantee full compensation it may lead to under-deterrence of the insured. From the moment of buying insurance, investment in preventive measures is no longer worthwhile for the insured, even if the cost is negligible. Consequently, liability insurance may reduce the incentive for efficient conduct provided by tort liability.[132] Although insurers may use various measures to encourage the insured to take precautions, such as deductibles, these measures cannot secure efficient conduct.[133] In addition, an insurer will often set a ceiling on coverage. If the insured's assets are insufficient to cover liability in excess of the cap, liability insurance will not solve the judgment-proof defendant problem.[134] In our context, while the extent of liability for defamation should not preclude full insurance coverage, liability insurance will clearly impair deterrence. Combining the two effects, we can expect amplification of the social imbalance of power: exclusion of weaker participants on the one hand and under-deterrence of stronger participants on the other hand.

A fourth possible solution is to impose criminal liability on those who are not induced to take optimal care by civil liability.[135] However, criminal law does not generally, nor is it supposed to, deal with non-intentional conduct (subject to limited exceptions), and should be used very sparingly to regulate free speech. In our view, it is inconceivable that while criminal

---

[128] *Id* at 169.

[129] *Id*; see also *Logue* (fn 122) 1375–1376.

[130] Electronic Frontier Foundation (EFF), Legal Guide for Bloggers, <https://www.eff.org/issues/bloggers/legal/liability/defamation>.

[131] Eugene Vologh, Bloggers—You Might Have Already Had Libel Insurance <http://www.volokh.com/posts/1185312054.shtml>.

[132] *I Englard*, The Philosophy of Tort Law (1993) 41–42; *I Englard*, The System Builders: A Critical Appraisal of Modern American Tort Theory (1980) 9 Journal of Legal Studies 27, 46; *JG Fleming*, Is There a Future for Tort? (1984) 44 Louisiana Law Review 1193, 1197 ('[T]he admonitory effect of an adverse judgment is today largely diffused by liability insurance which protects the injurer from having to pay the accident cost.'); *SD Sugarman*, Dowing Away with Tort Law (1985) 73 California Law Review 555, 574–582.

[133] *GT Schwartz*, The Ethics and the Economics of Tort Liability Insurance (1990) 75 Cornell Law Review 312, 337-338.

[134] *Logue* (fn 122) 1376, 1384.

[135] *Shavell* (fn 115) 170.

law experts already lament the excessive expansion of criminal liability[136] economic analysts of tort law would suggest expanding it even more.

4 Transaction costs

Now assume that the primary cyber-wrongdoer can be easily identified and is not judgment-proof. Imposing liability on him or her might still be economically unwarranted if a cheaper cost avoider exists. Assume, for example, that S makes a defamatory statement about P on C's online platform. There is a 10% chance that P will incur a €1000 harm following the publication, making the expected harm €100. S could determine that the statement was defamatory and false by taking precautions at a cost of €50. At this point it seems efficient to impose liability on S for the harm caused, because the expected harm was higher than the cost of avoidance (100>50). However, efficiency is determined not only by comparing the cost of precautions taken by S with the consequent decrease in expected harm, but also by comparing the cost of these precautions with those that can be taken by other parties with a similar effect. Thus, if C, as an experienced platform operator, could examine the statement at a cost of €30, C rather than S should be incentivised to take precautions. Imposing liability on S would be inefficient to the extent that it will induce S to take the more expensive precautions (€50>€30).

In accordance with the Coase theorem,[137] if externalities can be traded, and transaction costs are sufficiently low, an efficient outcome will ensue irrespective of the default legal rules.[138] In our example, if liability is imposed exclusively on the speaker S, he or she will pay C €30+x (where x<€20) to take the necessary precautions, saving S €20–x, and improving C's position by x. However, given the enormous number and high variance of the interactions between internet users and content providers transaction costs are very high. Determining who is the cheapest cost avoider for each potentially defamatory statement on the internet, and negotiating transfers from legal bearers of liability to cheapest cost avoiders, may be extremely complicated. Thus, where a content provider is the cheapest cost avoider but the law imposes liability only on the speaker, a contractual transfer is not guaranteed. We suspect, however, that this problem is not acute in our context because speakers do not seem systematically inferior cost avoiders compared to content providers.

[136] See, eg, *CB Ramsey*, Homicide on Holiday: Prosecutorial Discretion, Popular Culture, and the Boundaries of the Criminal Law (2003) 54 Hastings Law Journal 1641, 1642 & n 9 (observing that several legal scholars complain about the statutory expansion of the criminal law into areas that they believe should be regulated by tort).

[137] *RH Coase*, The Problem of Social Cost (1960) 3 Journal of Law & Economics 1.

[138] *Lichtman/Posner* (fn 108) 229-230.

C Exclusive indirect liability

1 The basic justifications

Content providers' liability is analogous to indirect liability in other contexts, such as employers' vicarious liability, bartenders' liability for harm caused by intoxicated patrons, motor vehicle owners' liability for harm caused by drivers to whom they loaned their cars, and landlords' liability for harm caused to tenants by trespassers.[139] In all cases the third party is capable of, and expected to, exercise control over the primary wrongdoer's conduct. Indirect liability might be required to secure efficient deterrence in two cases.

First, conventional economic analysis suggests that a rule imposing indirect liability is not necessary when the parties directly responsible for unlawful conduct can themselves be effectively deterred by legal sanctions.[140] As explained above, direct liability does not generate efficient deterrence if anonymous wrongdoers evade liability, either because of the high costs of identifying and pursuing them,[141] or because they are unable to fully compensate for the harm caused by their conduct.[142] Evasion of liability undermines the efficacy of defamation law, so that potential victims no longer have a meaningful protection of their reputation. A possible solution would be to impose liability on a third party who is (1) easily identifiable; (2) deep-pocketed; and (3) capable of controlling the primary wrongdoer's conduct. Content providers may often, though not always,[143] satisfy these three conditions. They are the gateway for online publication and therefore can prevent defamatory statements; in the event that they are business or government entities they are also identifiable and solvent.[144] Imposing indirect liability in such cases incentivises content providers to exercise their effective control over user-generated content, thereby preventing illegitimate speech. Put differently, indirect liability induces the content provider to prevent misconduct by potential wrongdoers.[145]

Second, indirect liability may be economically justified even where the speaker is easily identifiable and not judgment-proof, and hence sufficiently deterred by direct liability, if the content provider can prevent the expected harm at a lower cost than the speaker. As explained earlier, if transaction costs were low, speakers could shift the burden to content providers through agreement, and there would be no need for a rule of indirect liability. [146] However, because transaction costs are usually

---

[139] *Id* at 228.
[140] *Arrow et al* (fn 116) 4.
[141] Cf *Arrow et al* (fn 116) 2-3.
[142] *Lichtman/Posner* (fn 108) 229.
[143] Some content providers may be individuals who operate online forums, or the like, and have a limited capacity to compensate.
[144] *Lichtman/Landes* (fn 114) 398-399.
[145] *Hamdani* (fn 6) 912; *Lichtman/Landes* (fn 114) 398.
[146] *Lichtman/Posner* (fn 108) 229.

substantial, the burden will not be voluntarily shifted to the cheapest cost avoider. If content providers know that they are immune from liability as long as they provide a platform for a sufficient amount of legitimate speech, they will have no incentive to discourage illegitimate speech even if the costs of doing so are very low.[147] Therefore, content providers should be legally liable where they are the cheapest cost avoiders.

However, in each of these two cases one must examine the overall costs and benefits of imposing indirect liability. While direct liability may not secure efficient deterrence due to enforcement problems, indirect liability might not be economically justified. The analysis should take into account not only the standard variables, namely the cost of precautions and the reduction in expected harm, but also the administrative costs and other negative effects of liability.

Two additional justifications for indirect liability are unrelated to deterrence. The first concerns loss spreading. According to conventional economic wisdom, loss spreading reduces loss of welfare associated with risk aversion.[148] Thus, if we have to choose between two potential bearers of a particular loss, choosing the better loss spreader yields some economic benefit. Content providers can better spread costs through pricing and insurance than ordinary users. The second justification, which is unrelated to efficiency and mentioned here for the sake of completeness, focuses on distributive justice. Content providers, as opposed to the archetypal internet user, generate profit from operating online platforms. They can collect and commercialise user data, post advertisements, take part in marketing endeavors, etc. Because content providers profit from operating an online platform, they should be held accountable for at least some of the harmful outcomes of their activity.

2 The cost of precaution

a Types of precaution

When indirect liability is imposed on content providers they are incentivised to take precautions to reduce expected harm hence expected liability until the marginal cost of precaution exceeds the marginal reduction in expected liability. What kinds of precautions are available to content providers? First, they can screen user-generated content for defamatory statements and remove them when found. This can be done manually, by qualified and trained personnel, or automatically using natural language processing algorithms.[149] Second, content providers can employ a complaint/report management system, enabling affected parties or other users to report defamatory statement, examine complaints, and respond accordingly. Third, content providers can collect data on user activity,

[147] *Arrow et al* (fn 116) 9

[148] *R Perry*, Relational Economic Loss: An Integrated Economic Justification for the Exclusionary Rule (2004) 56 Rutgers Law Review 711, 758-761.

[149] *Lichtman/Posner* (fn 108) 230, 236.

based on IP addresses or registered user-names, and take measures against repeated wrongdoers, such as blocking. If the cost of these precautions is too high, indirect liability may, in some settings, be less cost-effective than direct liability. And if some social benefits are neglected by the content provider, greater costs of precaution lead to greater over-deterrence.

b Monitoring

Economists assume that the costs of the measures content providers can take to prevent wrongful speech are relatively low.[150] But this is not necessarily so. Let us begin with monitoring. Content providers can employ two types of monitoring systems—human or automated. The first type entails hiring and training people to read user generated content and distinguish between legitimate and non-legitimate speech. The cost per-statement is clearly substantial. Moreover, under an exclusive direct liability regime, the arguably prohibitive cost of identifying an anonymous speaker should be incurred only in the case of a defamatory statement that its victim wishes to bring to court. Under an exclusive indirect liability regime the cost of human supervision is incurred with respect to each and every statement: content providers cannot distinguish between legitimate and illegitimate speech without going through all content. For instance, assume online journal users make a thousand comments daily, only two are defamatory and only one of the two victims wishes to sue. Imposing direct liability would require identifying the author of a single comment, whereas imposing indirect liability would require examination of a thousand comments. This point cannot be understated. Assuming that most statements are non-defamatory, and that not all defamatory statements result in a lawsuit, the ratio between the number of statements and the number of claims may be extremely high. Thus, while content providers' supervision cost per-statement may be smaller than the cost of identifying an anonymous speaker in order to sue, the former must be multiplied by the large number of statements, most of which will not result in a lawsuit. The amount of data makes monitoring very costly.[151] Note further that the cost of human supervision is correlated with the amount of user-generated content, so the marginal cost may still be high. Using 'machine learning' algorithms may somewhat reduce this cost.

The second type of monitoring, namely an automated system, requires development and implementation of technologies which preclude defamatory statements while allowing legitimate speech. Presumably, this is a less expensive method than human monitoring because once the mechanism has been developed it can be implemented at a very low marginal cost. But automated systems are still expected to make more judgment-mistakes than trained humans, and given the asymmetric

[150] *Hamdani* (fn 6) 911; *Lichtman/Landes* (fn 114) 396

[151] *N Elkin-Koren*, Copyright Law and Social Dialogue on the Informative Superhighway: The Case against Copyright Liability of Bulletin Board Operators (1995) 13 Cardozo Arts & Entertainment Law Journal 345, 405.

treatment of errors discussed below, may be designed in a way that carries a greater risk of 'false positive' findings of defamation. A human-based correction system may be helpful, but it brings back at least some of the costs of human monitoring.

A possible response to the fear of high investigation costs is that if the costs of additional precautions (more monitoring) exceed the reduction in expected liability, content providers will simply not take these precautions; only if the cost of additional precautions is lower than expected reduction in liability will content providers employ them.[152] If both monitoring costs and expected liability are higher than the content provider's benefit from operating the platform, it will avoid the activity. Changes in levels of care or activity seem to ensue when they are efficient. So the fact that monitoring may be costly does not in itself justify exclusion of indirect liability. But this response ignores at least two relevant concerns. First, if the cost of content providers' precautions is indeed very high, exclusive direct liability might be preferable, despite its problems, to indirect liability. Second, in choosing the level of care and the level of activity the content provider might not capture relevant social benefits, so greater monitoring costs will result in greater over-deterrence. We will address this concern below.

c Notice and takedown

The second precautionary measure that content providers may employ is some form a 'notice and takedown' mechanism, namely removing user generated content when notified that this content is suspected of being defamatory. Only in the event that the content provider does not respond can it be found liable. This system is akin to the notice-and-takedown 'safe harbor' in the Digital Millennium Copyright Act (DMCA), which applies to online copyright infringements in the US.[153] Under the DMCA, if a copyright owner informs the content provider of an infringement, the latter must act.[154]

The main advantage of this method is that it reduces the need for monitoring significantly. The content provider does not need to examine each and every contribution, and must respond only in the case of complaint. The main cost of indirect liability, namely the cost of comprehensive monitoring, is saved. At the same time, the content provider's control over published material is utilised to prevent users from making defamatory statements, thereby enforcing the balance achieved within defamation law. Moreover, this method enables anonymous speech, with all the benefits it carries. Users do not need to provide personal information to contribute content.

In discussing the flaws of this method, we begin with its simplest version in which allegedly defamatory content is automatically removed

[152] *Hamdani* (fn 6) 915.
[153] 17 US Code § 512 (2000).
[154] *Lichtman/Landes* (fn 114) 402.

upon notification. Assuming people wish to silence their adversaries or competitors, they will notify content providers of any posting that is inconsistent with their views or interests. Put differently, anyone with the desire to silence another's speech will be able to do so easily, and to engage in mass censorship.[155] By analogy, empirical evidence indicates that more than a quarter of DMCA takedown notices are either on shaky legal grounds or address cases in which no copyrights are violated.[156] This may happen in the anonymous defamation context as well and is considered a plague on Facebook, a social network platform which has voluntarily implemented a notice-takedown policy. The price in terms of impairing freedom of speech and lost content is considerable because content providers take content out without verifying whether or not it is defamatory. [157] This is especially crucial for websites which provide extremely valuable content services based on low-cost administration, such as Craigslist.[158]

The problem of excessive limitation of free speech arising from automatic removal upon notice can be addressed in several ways, such as (1) requiring an exercise of human discretion prior to removal; (2) implementing an automatic content screening algorithm to certify the complaint; (3) allowing a counter-notice by the speaker; and (4) penalising complainants in cases of false accusations. The last two features are incorporated in the DMCA, which applies in a related context—online copyright infringements.

Human discretion will prevent, at least in part, removal of legitimate content following false accusations. But if human judgment is required, the marginal cost of handling an individual posting can be relatively high. The more postings, the higher the aggregate cost. Even if a content provider employs 'an army of lawyers,' it cannot adequately tell what content should legally stay up and which should be removed. [159] And if the cost of human investigation of complaints becomes prohibitive, the content provider may disable user contribution.

The court in *Zeran* explained that if the law imposed liability on content providers for not removing defamatory statements after receiving proper notices (as in the case of 'distributors' under the common law), then '[e]ach notification would require a careful yet rapid investigation of the circumstances surrounding the posted information, a legal judgment concerning the information's defamatory character, and an on-the-spot editorial decision whether to risk liability by allowing the continued publication of that information.'[160] The court added that '[a]lthough this

---

[155] *C Ziniti*, Note, The Optimal Liability System for Online Service Providers: How Zeran v. America Online Got It Right and Web 2.0 Proves It (2008) 23 Berkeley Technology Law Journal 583, 606.

[156] *Id* at 605.

[157] *Id* at 604.

[158] *Id*.

[159] *Id* at 607.

[160] *Zeran v America Online, Inc*, 129 F 3d 327, 333 (4th Cir 1997).

might be feasible for the traditional print publisher, the sheer number of postings on interactive computer services would create an impossible burden in the Internet context.'[161] Empirical data supports this concern.[162]

Arguably, requiring complainants to certify their legal claims and imposing criminal or civil liability in cases of false allegations (as anti-SLAPP laws often do)[163] could help avoid an overly restrictive outcome.[164] The same is true for setting up a counter-notification system, enabling speakers to challenge allegations of defamation. But these two mechanisms entail significant administrative costs which may also result in disabling user contribution. Moreover, neither can be effective unless someone is willing to contend that an accusation was false, and if the removed statement was made anonymously, the speaker may be unwilling to come forward. Thus, legitimate content will still be lost following uncontested complaints.

d Data collecting

Let us now turn to data collecting. On the technological level, keeping and processing user-activity logs require an increasing amount of memory and processing power over time. These resources are far from being costless. However, assuming that the platform operator can utilise the collected data, more data may yield a greater benefit. On the legal level, user-privacy might be jeopardised.[165] Recall that a short while ago the ECJ held that the Data Retention Directive, which required telecom companies to store user data for up to two years, was invalid for this reason. Lawmakers must find a way to encourage data collection and processing that is neither too difficult nor legally problematic. As we shall see below, data collection and processing may be more valuable in combined direct- and indirect-liability regimes.

3 Unaccounted benefits

In the previous section we argued that the cost of precautions taken by content providers may be relatively high. We explained, however, that this fact does not in itself justify exclusion of indirect liability. Yet the incentive structure in the online speech context may be tilted. Imposing liability on content providers makes them internalise the negative externalities of their activity generated by others under content providers' effective control. However, these indirect defendants provide platforms not only for defamatory statements but also, and mostly, for legitimate and valuable exchange of information and ideas. Web 2.0 users 'create positive

[161] *Id*.

[162] For example, Google reports millions of requests per month from government agencies, courts, and copyright owners, to remove content. See Google Transparency Report <https://www.google.com/transparencyreport/removals/>.

[163] State Anti-SLAPP Laws <http://www.anti-slapp.org/your-states-free-speech-protection/>.

[164] *Ziniti* (fn 155) 606.

[165] *Cf Arrow et al* (fn 116) 6.

externalities enjoyed by advertisers, information providers, merchants, friends, and acquaintances.'[166] The benefits of legitimate uses of the platform, not only the operator's gain, must be weighed against the costs of illegitimate uses. Only if one internalises the costs and benefits of one's conduct one would take socially optimal measures. Alas, under an indirect liability regime content providers internalise social costs without internalising all social benefits.[167] Because content providers do not capture the full social benefit of their activity, bearing the costs may result in overdeterrence—excessive monitoring and overzealous censorship.[168]

We illustrate this point using a simple numerical example based on Assaf Hamdani's seminal work on the subject.[169] Assume that the average public value of each statement made on an online platform is €10, whether or not defamatory. There is a 10% probability that a statement is defamatory, and the extent of harm caused by a defamatory statement is €50. The content provider can take one of three levels of care: (1) no monitoring ('level 0'); (2) monitoring at a cost of €1 per statement, reducing the likelihood of defamatory statements by 50%, that is, from 10% to 5% ('level 1'); or (3) monitoring at a cost of €1.2 per statement, reducing the likelihood of defamatory statements by 60%, that is, from 10% to 4%, but also erroneously identifying 10% of legitimate statements as defamatory ('level 2'). Table 1 shows the change in social welfare for each level of care per statement. 'Level 1' is the optimal level of care.

Level of care	Cost of Prec-aution	Expected harm	Expected benefit for the public	Total change
0	0	10%×50=5	10	10–5=5
1	1	5%×50=2.5	10–5%×10=9.5	9.5–1–2.5=6
2	1.2	4%×50=2	10–6%×10 –10%×90%×10= 8.5	8.5–1.2–2=5.3

Now assume that the law imposes indirect liability on the content provider whenever a defamatory statement is published on its online platform. Table 2 shows the content provider's private cost per statement under a strict liability rule. Recall that indirect liability, such as vicarious liability, is usually strict, even where the primary tortfeasor's liability is fault-based. A rational content provider would choose level 2, minimising

[166] *Lichtman/Posner* (fn 108) 225.
[167] *Lichtman/Posner* (fn 108) 238-239.
[168] *Hamdani* (fn 6) 916-918, 921; cf *N Netanel*, Impose a Noncommercial Use Levy to Allow Free Peer-to-Peer File Sharing (2003) 17 Harvard Journal of Law & Technology 1, 13 n 30 ('ISPs do not fully share the benefits its subscribers derive from placing material, whether infringing or non-infringing, on the network. As a result, imposing liability on ISPs for subscribers' infringing material induces ISPs to overdeter, purging any material that a copyright holder claims is infringing').
[169] *Hamdani* (fn 6) 919-920.

its private cost. In other words, the fact that the content provider does not capture the public benefit of published statements results in over-deterrence.

Level of care	Cost of prec-aution	Expected harm→ Liability	Content provider's private cost
0	0	10%×50=5	5
1	1	5%×50=2.5	1+2.5=3.5
2	1.2	4%×50=2	1.2+2=3.2

We have demonstrated that the content provider does not internalise the full social value of operating an open online platform. Indirect liability results in over-deterrence. As explained above, in a world of very low transaction costs, over-deterrence can be prevented through voluntary transactions. Those who benefit from online statements could pay content providers to choose the socially desirable level of care. In the above example, the expected public benefit is €9.5 per statement when the content provider chooses level 1, and €8.5 when it chooses level 2. So the public would pay up €1 to convince the content provider to choose level 1. The content provider's cost is €3.5 if it chooses level 1, and €3.2 if it chooses level 2. It will choose level 1 only if compensated for the additional cost (3.5−3.2=0.3). The contractual price would be x satisfying $0.3<x<1$ per statement.[170]

However, this solution is unrealistic, because transaction costs, particularly in the current context, are extremely high. Hamdani explains that there are at least three types of obstacles. First, the ordinary costs of negotiating and drafting an agreement.[171] Second, the benefit of online speech is widely disseminated. Perhaps users' aggregate benefit is significant, but a single user's benefit (a fraction of the total) may be too low to induce negotiations with the content provider.[172] Theoretically, users can negotiate as a group, but this raises collective action problems. Third, to enforce an agreement with respect to monitoring users need to obtain information about content providers' actual practice, and to prove in court that it differs from the agreed.[173] Enforcement costs greatly exceed a single user's expected gain, and any attempt for enforcement by a group of users might once again suffer from collective action problems.

Some believe that market forces (particularly dynamics of demand) can fix the incentive deficiency. Content providers with different monitoring levels will compete with each other. Users will move away from overly restrictive platforms, leading to less income for their operators from providing user-services, advertising, marketing projects, etc.[174] However,

[170] *Id* at 923.
[171] *Id* at 924.
[172] *Id* at 924-925.
[173] *Id* at 925.
[174] *Lichtman/Landes* (fn 114) 406; but see *Hamdani* (fn 6) 928 (criticising this view).

the likelihood of success is doubtful. It will be very difficult for users to determine which content providers are less restrictive. A promise is insufficient, because users cannot easily discern whether promises are kept. Content providers can try to establish general reputation of being 'less restrictive' to attract users, but this may be difficult to do, and increases the risk of liability.[175] A possible solution might be regulatory or statutory transparency requirements. More importantly, content providers compete on various levels, some — such as user interface or the scope and quality of the content — seem more important to users than monitoring level. Conceivably, users' choices will not be affected by monitoring practices. Lastly, due to powerful network effects a very few platforms tend to dominate the market (consider Facebook, for example), and vendor lock-in, that is, consumers' inability to replace their vendor without high switching costs (consider, again, Facebook), is a common phenomenon. In such instances one cannot rely on the market.

A third possible solution, proposed by Lichtman and Landes, is to subject content providers to fault-based rather than strict liability when a defamatory statement is published on their online platforms.[176] In deciding which level of care is socially desirable courts will take into account not only negative externalities but also positive externalities. Put differently, the court will set the standard of care in light of a comprehensive cost-benefit analysis, and content providers will not be liable if they monitor at the socially desirable level.[177] However, in our context, applying a fault-based indirect liability rule might be an extremely thorny if not an impossible task. It is very difficult to identify and evaluate the positive externalities of legitimate online publications—which are usually intangible, highly varied, and widely spread. Without this information courts cannot determine the socially desirable level of care. Moreover, to enforce a fault-based regime, victims need to obtain information about content providers' actual level of care, and then establish in court that it deviates from the socially desirable level. Furthermore, fault-based liability regimes introduce additional uncertainties which may result in over-deterrence.[178]

4 Asymmetric response to errors

Now assume that the content provider chooses the proper level of care. Some uncertainty may still arise with respect to the 'defamatoriness' (defamatory nature) of each statement. Very often it is difficult even for a qualified judge to determine whether a particular statement was defamatory. A content provider cannot reasonably employ a quasi-judicial system to examine all user-generated content. As explained above, it can employ human evaluators, but the cost of reducing the likelihood of errors

[175] *Hamdani* (fn 6) 929.
[176] *Lichtman/Landes* (fn 114) 405.
[177] *Hamdani* (fn 6) 934-935.
[178] See eg *R Perry*, Re-Torts (2008) 59 Alabama Law Review 987, 1003.

in terms of hiring more qualified personnel and training them may be exorbitant, and errors might occur even at the highest level of inspection. Presumably, automated systems are currently less capable of distinguishing legitimate and illegitimate speech.

These uncertainties force content providers to choose between two types of potential error: (1) a false negative, namely identifying a defamatory statement as non-defamatory and (2) a false positive, namely identifying a non-defamatory statement as defamatory. The problem hinges on the asymmetric legal response to each type of error: because a content provider's liability derives from the publication of a defamatory statement by a user, a false negative carries the risks of litigation and liability whereas a false positive does not. In our case, acting on a false positive does not seem to have a real cost at all (because removal is almost costless). This imbalance induces content providers to remove suspicious yet non-defamatory speech: to avoid liability, companies would rather err on silencing speech. The court in *Zeran* recognised this problem:[179]

> Because service providers would be subject to liability only for the publication of information, and not for its removal, they would have a natural incentive simply to remove messages upon notification, whether the contents were defamatory or not. Thus… [indirect] liability has a chilling effect on the freedom of Internet speech.

The problem arises not only in the case of monitoring but also in the case of employing a notice-and-takedown system: in unclear cases, rather than expend resources investigating incoming claims, interactive content providers will likely just take down the content.[180] Moreover, while automatic screening algorithms might be a lot cheaper to use, they are also expected to be overly restrictive given the asymmetric response to errors.

The asymmetric incentive structure may result in even harsher interference with speech-related interests. For example, content providers might identify risky users — those who publish 'grey area' statements, and block them, thereby giving up valuable contribution. As Neal Katyal explained: 'Because a [content provider] derives little utility from providing access to a risky subscriber, a legal regime that places liability on a [content provider] for the acts of its [users] will quickly lead the [content provider] to purge risky ones from its system.'[181] Moreover, when too many statements are provocative or suspicious, even if non-defamatory, the content provider may be led to disable user contribution of content saving substantial supervision costs.[182] Thus, 'the Internet might be about where digital cable systems are, with lots of downstream content and very little

---

[179] *Zeran v America Online, Inc*, 129 F 3d 327, 334 (4th Cir 1997).

[180] *Ziniti* (fn 155) 605-606.

[181] Neal Kumar Katyal, *Criminal Law in Cyberspace*, 149 U. PA. L. REV. 1003, 1007-08 (2001).

[182] *Ziniti* (fn 155) 600.

opportunity for interactivity, much less individual publishing.'[183] Finally, the fear of liability for user-generated content may impede technological innovation and development. Demand for new features that bring external content into play will decrease, and so will the incentive to develop such features.[184]

Online platforms may be used for defamatory and non-defamatory speech. The latter use is presumably much more frequent. The law should not interfere with the freedom of speech beyond the delicate balance achieved in defamation law between this freedom and the right to reputation. It should avoid significant interference with legitimate speech.[185] Yet removing borderline statements, blocking risky users, disabling user contribution, and reducing investment in developing interactivity tools — all serious threats to legitimate speech — are probable consequences of the asymmetric incentive structure.

A possible solution may be to create a counter-incentive, one which addresses the possibility of false positives. For example, the law can oblige content providers — under a threat of liability — to inform speakers of any block or removal of content and to enable them to challenge the decision. If a speaker challenges a decision, the content provider may inform the victim about the statement, enabling him or her to obtain an injunction. However, an attempt to remove the asymmetry by addressing false positives raises new problems. First, any mechanism based on communication between content providers and suspicious users requires collection of private information, and might jeopardise anonymity and infringe privacy. Second, and closely related, if contesting a restrictive decision requires disclosure of the speaker's identity, anonymous users, particularly those using constant anonymous profiles, may avoid it. This may result in loss of valuable content generated by anonymous users. Third, handling user objections entails significant administrative costs. Fourth, this particular method may not prevent a decision to disable user contribution and a reduction in demand for Web 2.0 features.

Arguably, the risk of disabling user contribution (the extreme consequence) may be somewhat reduced by market forces. If users prefer interactive web platforms, they may reveal their preferences by moving their activity from non-interactive to interactive platforms. The possible reduction in traffic may provide a sufficient incentive for enabling user contribution. Here, reliance on markets is more promising. It is not difficult for users to determine which platforms are interactive and which are not. However, the importance of interactivity for users depends on the nature of the platform. For example, in social networks interactivity is crucial,

---

[183] *J Harper*, Against ISP Liability (Spring 2005) 28 Regulation 30, 33 <http://ssrn.com/abstract=807545>.

[184] *Ziniti* (fn 155) 600.

[185] Cf *Arrow et al* (fn 116) 2-3, 6, 11 (making a similar explanation with respect to copyright violation); *Lichtman/Landes* (fn 114) 409; *Lichtman/Posner* (fn 108) 231-232.

whereas in online journal websites other aspects may affect users' preferences more than Web 2.0 features.

The American legislature has opted for an extreme solution. The Communications Decency Act discussed above granted immunity to internet service operators in order to 'promote the continued development of the Internet and other interactive computer services and other interactive media [and] to preserve the vibrant and competitive free market that presently exists for the Internet and other interactive computer services, unfettered by Federal or State regulation.'[186] It sought to prevent lawsuits from shutting down websites and other services which facilitate an exercise of the freedom of speech.[187]

D Concurrent liability

1 Advantages

A question arises whether de facto joint and several liability of the speaker and the content provider may be preferable to exclusive liability of either. The obvious advantage is that by imposing liability on content providers (in addition to online speakers) we can overcome the two main flaws of exclusive direct liability: the high cost of identifying and pursuing the anonymous speaker, and the problem of judgment-proof defendants. If the speaker can be identified at a very high cost or not at all, or if he or she cannot fully compensate the victim, indirect liability incentivises content providers to take the necessary precautions.

Concurrent liability also has another economic advantage. The interplay between two potentially liable parties generates an additional set of incentives. Parties who are jointly liable for a particular harm have an interest in reducing their own share of the burden. Because any difficulty in identifying and pursuing speakers will result in greater expected liability for the content provider, the latter has an incentive to facilitate identification of anonymous speakers. To do so, content providers may collect user information and volunteer this information in the case of a lawsuit.[188]

To begin with, the content provider can retain a log of users' IP addresses, disclosing those required to answer a claim for defamation. The cost increases with the number of users, and with the average user activity. However, this means is ineffective in cases of sophisticated wrongdoing, as where users use proxy servers or connect through public hot spots at cafés or the like. Alternatively, content providers may allow contribution only by registered users. One form of registration may enable anonymity even in the relation between a registered user and the content provider, but will

---

[186] 47 US Code § 230(b)(1)-(2).

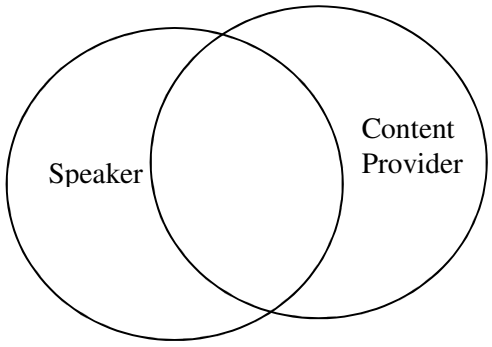
[187] *Batzel v Smith*, 333 F 3d 1018, 1028 (9th Cir 2002).

[188] Content providers might not be very keen to drag their users into court, because this may harm their business. But the ability to share the burden will surely result in some increase in the likelihood of data collection.

enable the latter to keep track of all IP addresses used by the former, and make evasion of liability less likely. A second form of registration will enable contribution through aliases, but require users to provide authentic personal information to the content provider upon registration. This path was taken by the South Korean legislature in the Real Name Verification Law of 2007, [189] but the Constitutional Court found this statute unconstitutional for violating the right to privacy and freedom of speech.[190] A third form of registration will allow contribution only under one's real name. The third option, which is the formal policy of Facebook, seems economically excessive because it eradicates online anonymity without any social benefit compared to the second option.

2 Disadvantages

Allowing both direct and indirect liability has several disadvantages. On the one hand, to the extent that both parties are at risk of being liable, and each has a somewhat different perception of what may constitute defamation, the actual limit on the freedom of speech may be excessive. The set of statements considered defamatory by either is the union of the set of statements considered defamatory by the speaker and the set of statements considered so by the content provider, which is equivalent to or larger than each (see figure below). Thus, imposing liability on both may restrict freedom of speech more than singling out one defendant ('double censorship'). On the other hand, imposing liability on each party reduces the other's incentive to take precautions.[191] The incentive may be diluted, leading to under-protection of reputation. Because these two conflicting effects cannot be evaluated, we are unable to ascertain whether and to what extent they cancel each other out.



Moreover, a combination of direct and indirect liability may result in an aggregation of the implementation costs of both. Content providers will

[189] *D Cho*, Real Name Verification Law on the Internet: A Poison or Cure for Privacy?, in: B Schneier (ed), Economics of Information Security and Privacy III (2013) 239.

[190] See *E Ramstad*, South Korea Court Knocks Down Online Real-Name Rule (24 August 2012) The Wall Street Journal <http://online.wsj.com/news/articles/SB10000872396390444082904577606794167615620>.

[191] *Lichtman/Posner* (fn 108) 239.

be led to monitor user-generated content at a high cost that could be saved under an effective direct liability regime. At the same time, lawsuits will be brought against speakers at a high administrative cost that could be saved under an effective indirect liability regime. Admittedly, the administrative cost of an action against the speaker may be lower than in the case of an exclusive direct liability regime due to the content provider's incentive to collect user information. But this reduced cost will add to the administrative cost of an action against the content provider, and to the costs of (excessive) monitoring.

Finally, as explained above, imposing liability on the speaker and the content provider incentivises the latter to collect data about the former. While reducing the cost of identifying and pursuing direct wrongdoers, this practice also raises privacy and data protection concerns. Databases compiled by content providers are subject to various risks, particularly security breaches and governmental abuse.

E Residual indirect liability

The last liability model, which we call 'residual indirect liability', is structured as follows: generally, the speaker alone will be liable for a defamatory online statement. However, if the speaker is unreachable, most likely because the content provider failed to collect user data and to provide it to the victim in the case of a lawsuit, indirect liability is imposed.[192] Thus, under the residual liability regime the content provider has to collect information that enables user identification in the case of an action for defamation, and to provide it to plaintiffs where such actions are brought. Indeed, some scholars advocate, and the British parliament endorsed, a system in which content providers' immunity hinges on their providing information about the violating speakers.[193]

At this point, after considering the alternatives, the advantages of this method should be evident. As opposed to exclusive indirect liability and concurrent liability, this liability regime may eliminate the need for monitoring, reducing content providers' monitoring costs to zero, and preventing over-deterrence associated with unaccounted benefits and asymmetric response to errors. Theoretically, if content providers under a residual liability regime allow postings by unreachable speakers, they might still need to monitor to avoid liability. Even so, monitoring will be limited to content generated by unidentifiable speakers, so the cost will be much lower than in the case of exclusive indirect liability or concurrent liability. More importantly, in practice, content providers would prefer to avoid liability through cheaper means, such as (1) obtaining user

[192] Meeting these two requirements grants some content providers immunity under sec 5 of the English Act.

[193] *P Ehrlich,* Note, Regulating Conduct on the Internet: Communications Decency Act § 230 (2002) 17 Berkeley Technology Law Journal 401, 402; *MA Lemley*, Rationalizing Internet Safe Harbors (2007) 6 Journal of Telecommunications and High Technology Law 101, 117-118 (discussing a system 'requiring intermediaries to retain and disclose the identity of their customers in response to a subpoena').

identification data, at least where an automatic content analysing algorithm finds the content potentially defamatory or (2) removing content generated by unreachable speakers upon notification of its defamatory potential. In fact, the new English Defamation Act incorporates both of these methods.

Additionally, the residual liability regime encourages content providers to collect user identification data, and facilitate actions against anonymous wrongdoers. This not only increases the likelihood of internalisation by speakers (hence efficient deterrence), but also reduces the administrative costs of claims against speakers. Finally, this liability regime does not raise the characteristic problems of multiple-defendants, such as excessive restriction of the freedom of speech on the one hand, and dilution of deterrence on the other hand. It also avoids the aggregation of the costs of two concurrent liability schemes. In summary, this method simultaneously solves the main problems of exclusive direct liability and exclusive indirect liability without the problems associated with multiple-defendants.

This method may raise difficulties at legal and economic levels, but the problems are arguably either insignificant or solvable. First, on the formal constitutional-legal level, the right to speak with anonymity may be a component of free speech.[194] By collecting and providing user information free speech is jeopardised. Furthermore, collecting user information infringes privacy. A possible response is that databases must be protected, and information may be disclosed only when a court determines that several preconditions—including, for example, a high likelihood that an action for defamation will succeed—are met. The English Act seems to provide an even better solution in allowing the speaker to decide whether he or she wishes to directly confront the victim or prefers that the statement be removed.

Second, on the economic level, two of the three problems associated with exclusive direct liability persist: content providers get off scot-free where the speaker is a judgment-proof defendant or when the content provider is the cheaper cost avoider. However, the probability that a content provider is a cheaper cost avoider than the speaker does not seem high. Additionally, although content providers are generally more deep-pocketed than users, the scope of the harm caused in typical online defamation cases may not be beyond the speaker's compensation capacity. The harm may be particularly small in the case of anonymous defamation, given the relatively weaker reliability and credibility of anonymous speakers. As explained above, the speaker may also have some insurance coverage. If there are settings in which speakers cannot normally

[194] *Doe v 2TheMart.com, Inc*, 140 F Supp 2d 1088, 1092 (D Wash 2001) ('A component of the First Amendment is the right to speak with anonymity. This component of free speech is well established... The right to speak anonymously extends to speech via the Internet.'); *ACLU v Miller*, 977 F Supp 1228, 1230-1232 (ND Ga 1997) (recognising the right to speak anonymously on the internet); see generally *McIntyre v Ohio Elections Commission*, 514 US 334 (1995) (the decision to remain anonymous is protected by the First Amendment).

compensate for the harm caused, an extension of content providers' liability under the residual liability regime may be warranted.

IV. CONCLUSION

This article examined various combinations of the two components of the legal response to online anonymous defamation—the ability or inability to identify and pursue an action against the anonymous speaker, and the legal recognition of content providers' liability for user-generated defamation. The Article analysed these combinations from comparative and economic perspectives.

The comparative analysis revealed four main paradigms for combining the two components. The US model bars content providers' indirect liability, but facilitates identification of the speaker. The Israeli model recognises content providers' fault-based liability but does not provide procedural tools for identifying the speaker. The EU framework enables the victim to request identification of the speaker, and at the same time bring an action against the content provider. Although there is variance among Member States, this model seems to comply with the relevant Directives and European court decisions. The recently-adopted English model enables the victim to pursue a claim against the speaker and, if the speaker is unavailable, imposes liability on the content provider. We called this 'residual indirect liability.'

From an economic perspective, the main problem with exclusively direct liability is that an action for online anonymous defamation entails special effort in identifying and pursuing the speaker. If the speaker is not identified, the costs of defamation are not fully internalised, and potential wrongdoers are not efficiently deterred. If the speaker is identified through a costly process, wrongdoers internalise the costs of their wrongdoing, but the administrative cost may outweigh the benefit in terms of cost-reducing deterrence. Alternatively, the high administrative costs associated with identifying the primary wrongdoer might render another party a more cost-effective target for enforcement efforts. An additional, yet probably less serious, problem marring the effectiveness of the incentives provided by direct liability is the high likelihood of judgment-proof defendants. Lastly, where a content provider is the cheapest cost avoider but the law imposes liability only on the speaker, a contractual transfer is not guaranteed due to high transaction costs. This problem too does not seem acute in our context.

The basic problem with exclusively indirect liability is the relatively high cost of precautions. Monitoring every user-generated statement is a very costly precaution. Notice-and-takedown is less costly, but an automatic system may result in excessive limitation of the freedom of speech, and human discretion once again entails very high costs, especially for websites with heavy traffic. Another problem is that content providers do not capture the full social benefit of their activity, so bearing the costs may result in overdeterrence—excessive monitoring and overzealous

censorship. A third problem is the asymmetric legal response to errors with respect to 'defamatoriness': because a content provider's liability derives from the publication of a defamatory statement by a user, a false negative carries the risks of litigation and liability whereas a false positive does not. This asymmetry induces content providers to limit speech-related interests.

Concurrent liability overcomes two flaws of exclusive direct liability: the high cost of identifying and pursuing the anonymous speaker, and the problem of judgment-proof defendants. Moreover, it induces the content provider to facilitate identification of anonymous speakers to reduce its own expected burden thereby increasing the likelihood of internalisation by the primary wrongdoer. But concurrent liability has several disadvantages. First, it has potentially conflicting effects on deterrence. On the one hand, assuming that the speaker and the content provider have different perceptions of what constitutes defamation, concurrent liability may restrict freedom of speech more than singling out one defendant. On the other hand, imposing liability on one reduces the other's incentive to take precautions, leading to under-protection of reputation. Second, a combination of direct and indirect liability may result in an aggregation of the implementation costs of both. Content providers' monitoring costs will add to the high administrative costs of lawsuits against speakers. Third, the content provider's incentive to collect data about speakers raises privacy and data protection concerns.

The residual indirect liability regime has several advantages. First, it eliminates (or at least reduces significantly) the need for monitoring, and prevents over-deterrence associated with unaccounted benefits and asymmetric response to errors. Second, it incentivises content providers to reduce the cost of identifying anonymous wrongdoers. Third, it does not raise the characteristic problems of multiple-defendants, such as excessive restriction of the freedom of speech or aggregation of costs. This model may also raise difficulties on the legal and economic levels, but they seem to us either insignificant (the judgment-proof defendant problem) or solvable (the risks to free speech and privacy).

This brings us to the conclusion that the English model, possibly with minor modifications, is the most efficient among the four existing paradigms. However, this conclusion must be taken with a pinch of salt. Our prescriptive analysis used cost-benefit terminology and, while the economic insights are clearly relevant for policy and law makers, there may be other considerations which ought to be taken into account in the final analysis. The catalogue of relevant considerations and their relative weights is in the end culturally contingent.