THE IRONY OF THE SAFE HARBOR REGIME: THE DISINCENTIVIZING LAWS ON INTERMEDIARY LIABILITY IN THE PHILIPPINES*

*Shiela Marie L. Rabaya***

ABSTRACT

Social media platforms fall under the definition of Internet Service Providers ("ISP") in the Cybercrime Prevention Act ("CPA"). Section 20 of the Implementing Rules and Regulations of the CPA provides an immunity from liability for ISPs, subject to three enumerated exceptions. One of these is the actual knowledge exception, which imposes liability upon an ISP if it is aware of the unlawful facts or circumstances in relation to the material it provides access to. This exception is in line with the safe harbor regime of intermediary liability since liability is contingent on actual knowledge. This apparent safe harbor poses a legal and policy dilemma, given its disincentivizing effect to the efforts of ISPs in regulating the content that it hosts. It is more prudent for ISPs to allow all types of content to be uploaded on their platforms instead of filtering defamatory and illegal content. This dilemma was akin to the issue sought to be eradicated by Section 230 of the Communications Decency Act in the United States (U.S.). Adopting the general immunity under Section 230, however, has its own set of problems. Constitutional and contract principles should be utilized to provide a true safe harbor for ISPs who exert Good Samaritan efforts to regulate content posted in its platform. A legal regime similar to the Digital Millennium Copyright Act and the ideas forwarded by the Manila Principles seem to be the mechanism that can achieve the balance between freedom of expression and necessary regulation.

I. INTRODUCTION

It is often heard that the pen is mightier than the sword. In the golden age of the Internet, however, the pen is mightier than ever. The Internet is "the most participatory form of mass speech yet developed."[1] It is a network

---

* *Cite as* Shiela Marie L. Rabaya, *The Irony of the Safe Harbor Regime: The Disincentivizing Laws on Intermediary Liability in the Philippines*, 94 PHIL. L.J. 140, [page cited] (2021).

** J.D., University of the Philippines College of Law (2020); B.A. Psychology, *magna cum laude,* University of the Philippines (2016); Dean's Medalist for Academic Excellence;

that knows no bounds, and the World Wide Web is a "unique and wholly new medium of worldwide human communication."[2] Regulation is then up to the law. In fact, "the growing public awareness of the Internet's unwieldy and chaotic side has led to calls for regulation and governance."[3] The regulation, however, must be within the bounds of the constitutionally protected rights of the people.[4] The pen may be mighty, but the Constitution shall reign supreme.

The Internet allows people all over the world to communicate with each other. The reach of the Internet is massive, such that it can even be said to be the only global medium that is accessible anywhere on earth.[5] Worldwide communication is done through social media platforms such as Facebook, Twitter, Instagram, and YouTube. Conventionally, third parties serve as the content providers while the social media platforms perform the role of Internet Service Providers ("ISP"). The ease of access to content uploaded to the Internet leads to questions regarding the proper regulation for such content. Since third parties often provide content, it is impossible to foresee what types of content will be uploaded. This unpredictability makes the realm of the Internet harder to regulate, especially with regard to defamatory or illegal content uploaded by third persons. The increase in the capacity of individual third persons to communicate with one another is extremely prone to abuse.[6]

Prior to the advent of the Internet, traditional liability attaches to the speaker of a defamatory statement. Such defamatory statements are made through written or printed word and the author of such statements may be held liable for libel under the Revised Penal Code ("RPC").[7] With the rapid

---

Member, Order of the Purple Feather (2018-2020); Editor, Student Editorial Board, PHILIPPINE LAW JOURNAL Vol. 92; Editorial Assistant, PHILIPPINE LAW JOURNAL Vol. 91; Research Assistant, UP Law Center Technology, Law and Policy Program (2019). The author would like to thank her family and friends for their never-ending support.

The author has previously published an Essay on intermediary liability in the COVID-19 Special Online Feature of the PHILIPPINE LAW JOURNAL Vol. 93, focusing on the role of intermediaries in regulating COVID-19 false content in the Philippines. This Note, on the other hand, is an analysis of the disincentivizing nature of the safe harbor regime in Philippine intermediary liability laws.

[1] Reno v. American Civil Liberties Union, 521 U.S. 844, 869 (1997).

[2] *Id.* at 850.

[3] *Developments in the Law: The Law of Cyberspace*, 112 HARV. L. REV. 1574, 1581 (1999).

[4] CONST. art. III, § 4.

[5] Obiageli Ohiagu, *The Internet: The Medium of the Mass Media*, 16 KIABARA J. HUMAN. 225 (2011).

[6] Ronald Mann & Seth Belzley, *The Promise of Internet Intermediary Liability*, 47 WM. & MARY. L. REV. 239, 244–45 (2005).

[7] REV. PEN. CODE, art. 353.

growth of the Internet, there existed a gap as to the liability for defamatory statements made online. It was initially argued that defamatory statements made through the Internet are still punishable under the RPC.[8] The gap was later addressed, however, by the passage of the Cybercrime Prevention Act of 2012 ("CPA").[9]

The CPA was "enacted by Congress to address legitimate concerns about criminal behavior on the Internet and the effects of abusive behavior[.]"[10] Section 4(c)(4) provides the liability for libel as defined in the RPC but "committed through a computer system or any other similar means[,] which may be devised in the future."[11] The law then addresses the liability for defamatory statements made through the Internet by imposing liability upon the authors of such statements. The gaps in the law with regard to liability for defamatory statements committed on the Internet then seemed to have been resolved.

A suit filed in May 2019, however, sheds light on another facet of liability on the Internet. A Bicolano tycoon filed cyber libel cases against one Peter Joemel Advincula and social media giants, Facebook and Google Philippines. The suits were filed due to the defamatory statements made by Advincula in videos which became viral through the platforms of Facebook and YouTube (owned by Google Philippines). The statements allegedly implicated the tycoon in illegal drug trade dealings in the Philippines. Facebook and YouTube were sued because they allegedly allowed their online platforms to be used in spreading the false and malicious information.[12] Moreover, the media giants also allegedly refused to take down the videos even after a formal request by the Bicolano tycoon.[13] Applying the concept of traditional libel under the RPC *vis-à-vis* the CPA points to the liability of Advincula, if he is indeed found guilty of making such defamatory statements. The liability of Facebook and YouTube as "intermediaries," however, is not

---

[8] Oscar Franklin Tan, *Supreme Court Idol: The cyberlibel edition*, RAPPLER, Jan. 21, 2013, *at* https://www.rappler.com/voices/thought-leaders/supreme-court-idol-the-cyberlibel-edition.

[9] Rep. Act No. 10175 (2012).

[10] Kim Arveen Patria, *Palace: Thou shall not fear cybercrime law*, YAHOO! NEWS, Oct. 3, 2012, *at* http://ph.news.yahoo.com/palace--thou-shall-not-fear-cybercrime-law.html.

[11] Rep. Act No. 10175 (2012), § 4(c)(4).

[12] Rhaydz B. Barcia, *Bicolano tycoon sues Bikoy, Facebook, Youtube for cyber libel*, RAPPLER, May 23, 2019, *at* https://www.rappler.com/nation/231382-bicolano-tycoon-sues-bikoy-facebook-youtube-cyberlibel.

[13] Edu Punay, *P1 billion cyber libel suit filed vs Facebook, Youtube, 'Bikoy,'* PHILSTAR GLOBAL, May 24, 2019, *at* https://www.philstar.com/headlines/2019/05/24/1920450/p1-billion-cyber-libel-suit-filed-vs-facebook-youtube-bikoy.

as clear under current Philippine law and jurisprudence. The suit filed by the Bicolano tycoon is the first of such kind in Philippine jurisdiction.[14]

Questions now arise as to the liability of "intermediaries" such as Facebook and YouTube. Such intermediaries fall under the term "service provider" or ISPs as defined in the CPA and its Implementing Rules and Regulations ("IRR"). The said law and its IRR define a "service provider" as "any public or private entity that provides to users of its service the ability to communicate by means of a computer system" and "any other entity that processes or stores computer data on behalf of such communication service or users of such service."[15] A service provider which "willfully abets or aids in the commission of any of the offenses enumerated in [the] Act shall be held liable."[16] Liability is also imposed if service providers fail to preserve computer data within a specified period[17] or to disclose such traffic data and subscriber information after being compelled to do so by authorities.[18] The provisions on the liability of service providers seem straightforward in that the aiding and abetting must be done willfully in order to be punishable under the law.

The extent of the liability of a service provider, however, is given more detail in the IRR of the CPA. In particular, Section 20 provides the following:

> Section 20. *Extent of Liability of a Service Provider.* – Except as otherwise provided in this Section, *no person or party shall be subject to any civil or criminal liability* in respect of a computer data for which the person or party acting as a service provider *merely provides access* if such liability is founded on:
>
> a. The obligations and liabilities of the parties under a computer data;
> b. The making, publication, dissemination[,] or distribution of such computer data or any statement made in such computer data, including possible infringement of any right subsisting in or in relation to such computer data: Provided, That:
>
> 1. The service provider does not have actual knowledge, or is not aware of the facts or circumstances from which it is apparent, that the making, publication, dissemination[,] or distribution of such material is unlawful or infringes any rights subsisting in or in relation to such material;

---

[14] *Id.*

[15] Rep. Act No. 10175 (2012), § 3(n); Rep. Act No. 10175 (2012) Rules & Regs, § 3(ff).

[16] Rep. Act No. 10175 (2012), § 5(a).

[17] § 13.

[18] § 14.

> 2. The service provider does not knowingly receive a financial benefit directly attributable to the unlawful or infringing activity; and
> 3. The service provider does not directly commit any infringement or other unlawful act, does not induce or cause another person or party to commit any infringement or other unlawful act, and/or does not directly benefit financially from the infringing activity or unlawful act of another person or party[.][19]

Section 20 provides an immunity from liability for service providers, subject to the three enumerated exceptions in paragraph (b). This Note aims to illustrate and analyze the problems with the current legal regime on intermediary liability in the Philippines. The "actual knowledge" exception,[20] in particular, may pose a legal and policy dilemma.

The "actual knowledge" exception in Section 20(b)(1) requires a situation where the service provider is aware of the unlawful facts or circumstances in relation to the material it provides access to. If applied to the suit filed by the Bicolano tycoon, Facebook and Google Philippines may be held liable if it is proven that they were aware of the illegal nature of the videos and of the fact that they were posted on their platforms. Conversely, the ISPs would be immune from liability if they do not have such actual knowledge. It then becomes wise for ISPs to avoid actual knowledge at all costs, and this illustrates how the current legal regime disincentivizes regulation by ISPs. It is safer to allow all types of content to be uploaded on their platforms, instead of imposing regulations that filter defamatory and illegal content. Imposing such regulations may open ISPs to liability under the actual knowledge exception.

The aforementioned dilemma was precisely the issue sought to be eradicated by Section 230 of the Communications Decency Act ("CDA") in the U.S. Section 230 was passed after the decision in *Stratton Oakmont, Inc. v. Prodigy Services Co.*[21] highlighted a disincentive to self-regulation by ISPs in the legal regime prior to Section 230. This Note aims to show that the same disincentive is present in the current legal regime of intermediary liability in the Philippines. Part II, in particular, discusses the relevant legal concepts to intermediary liability. It will highlight the disincentivizing effect of the law prior to Section 230 of the CDA and how such problem was addressed by the passage of the latter law. Part III discusses Philippine laws on intermediary

---

[19] Rep. Act No. 10175 (2012) Rules & Regs, § 20. (Emphasis supplied.)

[20] § 20(b)(1).

[21] [Hereinafter "*Stratton Oakmont*"], 23 Media L. Rep. (BNA) 1794, 1796–98 (N.Y. Sup. Ct. 1995).

liability. In particular, the general constitutional policies behind intermediary liability and the current laws which branched out of such policies will be discussed. Part IV presents the author's analysis of the different regimes of intermediary liability in general and as applied in current Philippine laws. It will also explore options on how the disincentivizing effect of the current law can be avoided. Lastly, Part V presents the author's conclusion and recommendation.

II. RELEVANT LEGAL CONCEPTS TO INTERMEDIARY LIABILITY

With its unprecedented growth, the Internet has extended its reach to many different facets of society. It is also now infused with a variety of sociological and legal concepts.[22] Philippine intermediary laws have tried to keep up with the rapid changes on the Internet in order to safeguard society from its potentially harmful effects. In order to fully understand the nuances of intermediary liability for ISPs, it is useful to examine the different legal concepts that have developed in conjunction with intermediary liability.

A. Distinction Between Distributor and Publisher

In the U.S., the distinction between distributor and publisher was first seen as important in attaching liability to newspapers or other publications. A publisher is an entity who is responsible for the "creation or editing of content in a publication."[23] A distributor, on the other hand, is an entity "that makes publications available to the public."[24] Examples of a publisher are a book or newspaper publisher, while a bookseller or library are examples of a distributor.[25] Generally, the plaintiff need not prove that the defamation defendant was aware of the content of the defamatory statement if it is a *publisher*. Such proof, however, is required if the defamation defendant is a *distributor*.[26] During the early years of the Internet, there was no clear answer as to whether ISPs should be treated as a publisher or distributor in order to determine its liability for defamatory content in its platform. The enactment

---

[22] *See* Manuel Castells, *The Impact of the Internet on Society: A Global Perspective*, *in* CHANGE 127 (2014).

[23] David Sheridan, *Zeran v. AOL and the Effect of Section 230 of the Communications Decency Act Upon Liability for Defamation on the Internet*, 61 ALB. L. REV. 35, 150. *See Stratton Oakmont*, 23 Media L. Rep. (BNA) at 1796–98.

[24] Sheridan, *supra* note 23, at 150. *See also* Cubby, Inc. v. CompuServe, Inc. [hereinafter "*Cubby*"], 776 F. Supp. 135, 139 (S.D.N.Y. 1991).

[25] Sheridan, *supra* note 23.

[26] *Id. See also* Spence v. Flynt [hereinafter "*Spence*"], 647 F. Supp. 1266, 1273 (D. Wyo. 1986).

of Section 230 of the CDA, however, granted an "interactive computer service" immunity from liability either as a publisher or speaker of false and defamatory material.[27]

In fine, Section 230 provides that "[n]o provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider."[28] With the immunity provided by Section 230 of the CDA, it may be argued that the distinction between publisher and distributor became irrelevant. However, the only clear immunity provided by the wording of the law is that an ISP is immune from liability as a publisher. It is not clear whether ISPs are also immune from liability as a distributor. The "speaker" provided in the law does not necessarily equate to distributor. The Court in *Zeran v. America Online, Inc.*[29] decided that the immunity from publisher liability afforded to ISPs by Section 230 is extended to distributor liability. However, the Fourth Circuit in *Zeran* failed to consider the development of the common law definition of defamation, which made liability as distributor and as publisher distinctly different.[30] This issue remains unresolved until today, with *Zeran* remaining as the controlling interpretation.[31]

It is worthy to note, however, that Section 230 makes an important distinction between two types of cyber-entities. The distinction is between "interactive computer services" and "information content providers."[32] The statute provides a presumption that any participant in the entire Internet connection process is an interactive computer service.[33] However, if such participants create any information, they acquire the status of information content provider.[34] The immunity provided by Section 230 of the CDA is only for "interactive computer services." This type of cyber-entity cannot be treated as a publisher or speaker of a content provider's content. The distinction between "interactive computer services" and "information content

---

[27] *Id.* at 154.

[28] Communications Decency Act, 47 U.S.C., § 230(c)(1).

[29] [Hereinafter "*Zeran*"], 129 F.3d 327 (4th Cir. 1997).

[30] William Buelow III, *Re-Establishing Distributor Liability on the Internet: Recognizing the Applicability of Traditional Defamation Law to Section 230 of the Communications Decency Act of 1996*, 116 W. VA. L. REV. 313, 317–36 (2013).

[31] Ryan French, *Picking up the Pieces: Finding Unity after the Communications Decency Act Section 230 Jurisprudential Clash*, 72 LA. L. REV. 443, 454 (2012).

[32] *Id.* at 449.

[33] *Id. See* Communications Decency Act, 47 U.S.C. (2006), § 230(f)(2)–(3).

[34] Communications Decency Act, 47 U.S.C., § 230(f)(2).

providers" is one of content creation,[35] much like the distinction between publisher and distributor in earlier laws.

In the Philippines, the distinction between "publisher" and "distributor" can also be seen in the current legal regime of intermediary liability. Section 20 of the IRR of the CPA, which provides that no party acting as a service provider shall be held liable for computer data it "merely provides access" to, was substantially reproduced from Section 30 of the Electronic Commerce Act.[36] A service provider which "merely provides access" to third party content performs functions akin to that of a distributor which just "makes publications available to the public,"[37] or to "interactive computer services" of Section 230 of the CDA.

B. Regimes of Liability

There are three legal regimes governing intermediary liability. These are *strict liability*, *safe harbor*, and *general immunity*. Under the strict liability regime, intermediaries are liable in the same way that content providers are for illegal content. This regime holds the intermediary liable regardless of its knowledge and extent of control over the content disseminated through its platform. ISPs are then burdened with the duty to regulate the content on its platform to ensure that no illegal content is disseminated. The regime is usually applied in states where intermediaries have been used to propagate "subversive, seditious, and politically unsettling material."[38] The intermediaries then act as an arm of state censorship.[39] In most jurisdictions, however, the strict liability regime is seen as an impediment on the freedom of speech by creating a chilling effect.[40]

Under the safe harbor regime, intermediaries are only held liable for defamatory or illegal content if they had knowledge—actual or constructive[41]—that their platform contained such kind of content. [42] Actual knowledge refers to an intermediary's awareness and intention to violate the

---

[35] § 230(f)(3).

[36] Rep. Act No. 8792 (2000).

[37] Sheridan, *supra* note 23, at 150. *See also Cubby*, 776 F. Supp. 135, 139.

[38] Gemmo Fernandez & Raphael Lorenzo Pangalangan, *Spaces and Responsibilities: A Review of Foreign Laws and an Analysis of Philippine Laws on Intermediary Liability*, 89 PHIL. L.J. 761, 771–72 (2015).

[39] *Id., citing* Chris Reed, *Liability of Online Information Providers – Towards a Global Solution* 17 INTL. REV. L. COMPUT. & TECH. 255 (2003).

[40] *Id., citing* Benoît Frydman & Isabelle Rorive, *Regulating Internet Content through Intermediaries in Europe and the USA*, 23 ZEITSCHRIFT FUR RECHTSSOZIOLOGIE 41, 44 (2009).

[41] *Id., citing* Larusdottir, *infra* note 113, at 473.

[42] *Id.* at 772.

law by allowing defamatory or illegal content on its platform. There is constructive knowledge, on the other hand, if the intermediary should have reasonably presumed under certain factors that a material is illegal or infringing on an individual's rights.[43] Under constructive knowledge, there is no actual awareness or intention to include defamatory or illegal content on its platform. A "notice and takedown mechanism" is often included in safe harbor laws. This mechanism requires intermediaries to remove or disable access to illegal content upon receiving knowledge of its existence on the platform.[44]

Finally, the third regime is general immunity. This is the regime followed by Section 230 of the CDA, which provides that "[n]o provider or user of an interactive computer service shall be treated as publisher or speaker of any information provided by another content provider."[45] Unlike the safe harbor regime, there is no exception for actual or constructive knowledge. Under this regime, "intermediaries left to their own devices will, for commercial reasons, naturally take on an editorial and filtering role, so long as they are given protection from the risk entailed in being seen as publishers, distributors[,] or the like."[46] This is in line with the two primary reasons behind Section 230 of the CDA, which are to protect the ISPs who make the effort to regulate its content and to allow Internet companies to grow without the fear of crippling regulation.

C. Section 230 of the Communications Decency Act

The passage of Section 230 of the CDA was heavily influenced by the case of *Stratton Oakmont v. Prodigy Services Co.* In *Stratton,* a securities investment-banking firm sued a website operator named Prodigy Services Company for statements posted on Prodigy's "Money Talk" computer bulletin board. Prodigy was held liable for such defamatory remarks because it made efforts to filter inappropriate content, hence, it shouldered the burden of liability for any defamatory content that was not removed during its screening process. According to the Court: "Prodigy held itself out to the public and its members as controlling the content of its computer bulletin

---

[43] *Id.* at 773.

[44] *Id.*

[45] Communications Decency Act, 47 U.S.C. § 230.

[46] Fernandez & Pangalangan, *supra* note 38. *See* Charlotte Waelde & Lilian Edwards, Online Intermediaries and Copyright Liability (World Intellectual Property Organization Workshop Keynote Paper, Geneva) 2005.

boards."[47] The Court's analysis was then based on the editorial control Prodigy exercised over the content posted on the site. Prior to *Stratton*, courts refused to impose liability on websites that knew nothing of its content. With this reasoning, it is then better for ISPs to not regulate its content at all than to impose any screening measures. Failed attempts at screening content are punished far more harshly than not having any kind of screening measure. The disincentivizing nature of the legal regime prior to Section 230 is then apparent.

As a response to the *Stratton* dilemma, Section 230 of the CDA was passed in February 1996. The bill was sponsored by Representatives Christopher Cox and Ron Wyden.[48] Section 230 appears to immunize ISPs from two forms of liability:

> (1) [T]he inequitable Stratton dilemma, whereby a website could be held liable as the publisher of all information because of its attempt to filter some of the information; and (2) liability to those whose content a website filters, although the content is constitutionally protected. Roughly translated, *websites would not face liability for not blocking enough content or for blocking too much content.*[49]

To fully appreciate the intent behind Section 230, it is imperative to examine the speech of Congressman Cox during his discussion with the House of Representatives on August 4, 1995:

> We want to encourage people like Prodigy, like CompuServe, like American Online, like the new Microsoft network, to do everything possible for us, the customer, to help us control, at the portals of our computer, at the front door of our house, what comes in and what our children see. This technology is very quickly becoming available, and in fact everyone one [sic] of us will be able to tailor what we see to our own tastes.
>
> We can go much further, Mr. Chairman, than blocking obscenity or indecency, whatever that means in its loose interpretations. We can keep away from our children things not only prohibited by law, but prohibited by parents. That is where we should be headed, and that is what the gentleman from Oregon [Mr. Wyden] and I are doing.

---

[47] Andrew Bolson, *Flawed but Fixable: Section 230 of the Communications Decency Act at 20*, 42 RUTGERS COMPUTER & TECH. L.J. 1, 4 (2016).

[48] *Id.* at 5.

[49] French, *supra* note 31. (Emphasis supplied.)

> Mr. Chairman, our amendment will do two basic things: First, it will *protect computer Good Samaritans*, online service providers, anyone who provides a front end to the Internet, let us say, who takes steps to screen indecency and offensive material for their customers. It will protect them from taking on liability such as occurred in the *Prodigy* case in New York that they should not face for helping us and for helping us solve this problem. Second, it will establish as the policy of the United States that we do not wish to have content regulation by the Federal Government of what is on the Internet, that we do not wish to have a Federal Computer Commission with an army of bureaucrats regulating the Internet because frankly the Internet has grown up to be what it is without that kind of help from the Government. In this fashion we can encourage what is right now the most energetic technological revolution that any of us has ever witnessed. We can make it better. We make sure that it operates more quickly to solve our problem of keeping pornography away from our kids, keeping offensive material from our kids, and I am very excited about it.[50]

Section 230 was then intended to protect computer "Good Samaritans" or those ISPs who make the effort to regulate its content. Also, the statute was intended to allow Internet companies to grow without the fear of crippling regulation.[51] As a result, an ISP would not face intermediary liability whenever it chooses to edit or screen their platform for illegal content, or whenever an ISP had notice of such illegal content appearing on its platform.[52]

More importantly, Section 230 is vital in ensuring freedom of speech on the Internet.[53] Without the protection of Section 230, ISPs would most likely eliminate all interactive features and user-generated content, thereby limiting the platform of people for communication. The elimination of these fora for communication would have a direct detrimental effect to freedom of speech.[54]

---

[50] Bolson, *supra* note 47, at 7–8. (Emphasis supplied.) *See* 141 Cong. Rec. H8468 (daily ed. Aug. 4, 1995) (statement of Rep. Cox).

[51] Bolson, *supra* note 47.

[52] Andrew Sevanian, *Section 230 of the Communications Decency Act: A "Good Samaritan" Law Without the Requirement of Acting as a "Good Samaritan"*, 21 UCLA ENT. L. REV. 121, 125 (2014).

[53] Noah Tischler, *Free Speech under Siege: Why the Vitality of Modern Free Speech Hinges on the Survival of Section 230 of the Communications Decency Act*, 24 TEMP. POL. & CIV. RTS. L. REV. 277, 278 (2014).

[54] *Id.*

It became apparent later, however, that the sponsors of the bill were not able to anticipate the magnitude of the Internet's growth in the years following Section 230's passage. Since then, Section 230 has been heavily criticized due to the perceived blanket immunity provided by the statute to ISPs. Such blanket immunity was seen as a tool that enabled content "that bullies, harasses, intimidates, impersonates[,] and exploits children[,] and that neither parents nor anyone else would have the ability to control the dissemination of such content."[55] It would then seem that Section 230 was not able to keep up with the rapid changes in the cultural and societal atmosphere surrounding the Internet.

III. PHILIPPINE LAWS ON INTERMEDIARY LIABILITY

The Philippines does not share a similar trajectory of Internet legislation as that of the U.S. Defamatory statements were first punished through libel under the RPC. Later, the CPA and its IRR governed defamatory statements communicated through the Internet.

Prior to the advent of the Internet, newspapers and other similar publications served the role of intermediaries. Such publications, however, only served as a platform for a select pool of writers and the final output goes through a whole editorial process. This is why the distinction between distributor and publisher was crucial in determining the liability of such publications under the law on libel. Traditional liability attaches to the speaker or writer of a defamatory statement under Article 353 of the RPC. Article 353 defines libel as a "public and malicious imputation of a crime, or of a vice or defect, real or imaginary, or any act, omission, condition, status, or circumstance tending to cause the dishonor, discredit, or contempt of a natural or juridical person, or to blacken the memory of one who is dead."[56] Libel is made applicable to writings or similar means through Article 355, which provides that "libel committed by means of writing, printing, lithography, engraving, radio, phonograph, painting, theatrical exhibition, cinematographic exhibition, or any similar means, shall be punished by *prision correccional* in its minimum and medium periods or a fine[.]"[57] The liability of the platform for expression such as newspapers and other similar publications is dependent on the court's determination of whether the platform is a publisher or a distributor. Generally, the plaintiff need not prove that the defamation defendant was aware of the content of the defamatory statement

---

[55] Bolson, *supra* note 47, at 9.
[56] REV. PEN. CODE, art. 353.
[57] Art. 355.

if it is a publisher. Such proof, though, is required for liability to attach if the defamation defendant is a distributor.[58]

The birth of the Internet, however, led to an unprecedented ease in communications. An opinion can easily be expressed through online intermediaries or ISPs, and this opinion can reach anyone with access to the Internet. Unlike before where the speaker is held liable for his defamatory statements, it is now difficult to impose liability to the speaker because of the anonymous nature of the Internet.[59] The lines are now blurred between the sender and the receiver.[60] Intermediaries are thus an attractive target for legal claims, since these are seen as the "most effective point of control"[61] over internet-related misconduct.[62] Furthermore, the commercial interests that intermediaries have in content hosting and distribution strengthen the reasoning behind intermediary liability.[63] Lastly, targeting the distribution network of defamatory and other illegal statements is now the most strategic method of law enforcement.[64]

The effect of the imposition of liability on intermediaries depends on the regime of liability[65] followed by the regulatory law. Historical antecedents of the Internet suggest that imposition of strict intermediary liability threatens innovation and free expression.[66] What remains clear, however, is that

---

[58] Sheridan, *supra* note 23. A distributor may also be held liable in the rare case in which a plaintiff shows that the distributor was aware of the source of the defamatory utterance and was aware of prior false and defamatory utterances by the same source and, therefore, had reason to know that the utterance that is the subject of the lawsuit was likely to be false and defamatory. *See Spence*, 647 F. Supp. 1266, 1273.

[59] Seth F. Kreimer, *Censorship by Proxy: The First Amendment, Internet Intermediaries and the Problem of the Weakest Link*, 155 U. PA. L. REV. 11 (2006).

[60] American Civil Liberties Union v. Reno, 929 F. Supp 824, 843, 883 (Dist. Court ED Pa 1996).

[61] Kreimer, *supra* note 59.

[62] Ronald Mann & Seth Belzley, *The Promise of Internet Liability*, 47 WM. & MARY L. REV. 239, 265 (2005); David Kaye (Special Rapporteur), Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, UNHRC, 38th Sess., ¶ 14, UN Doc A/HRC/38/35 (2018); Delfi AS v. Estonia, Grand Chamber Judgment, App. No. 64569/09, ¶ 111 (ECtHR, June 16, 2015).

[63] Delfi AS v. Estonia, Grand Chamber Judgment, App. No. 64569/09 (ECtHR, June 16, 2015).

[64] New York v. Ferber 458 U.S. 759 (1982); Danielle Keats Citron & Mary Anne Franks, *Criminalizing Revenge Porn*, 49 WAKE FOREST L. REV. 345 (2014).

[65] The three regimes of liability are *strict liability*, *safe harbor*, and *general immunity*.

[66] Center for Democracy & Technology, *"Regardless of Frontiers:" The International Right to Freedom of Expression in the Digital Age* (April 2011), *at* https://cdt.org/wp-content/uploads/pdfs/CDT-Regardless_of_Frontiers_v0.5.pdf.

expression on the Internet is protected by the same fundamental rights that safeguard traditional speech and expression.

A. Freedom of Expression[67]

The fundamental rights of freedom of speech and expression are protected in several Philippine and international legal instruments. Article III, Section 4 of the Constitution provides that "[n]o law shall be passed abridging the freedom of speech, of expression, or of the press, or the right of the people peaceably to assemble and petition the government for redress of grievances."[68] Such article protects the dual aspects of freedom of expression, namely the freedom from censorship or prior restraint and the freedom from subsequent punishment.[69] Internationally, the Philippines voted in favor of the Universal Declaration of Human Rights ("UDHR")[70] and later on became a State party to the International Covenant on Civil and Political Rights ("ICCPR"),[71] both of which embody several principles pertaining to freedom of expression. While the former is admittedly not a treaty, it has nevertheless become a normative instrument that creates "legal and moral obligations for Member States of the UN."[72]

Within the context of cyberspace, Article 19 of the UDHR appears to be applicable to expression on the Internet, given that it includes the words "through any media."[73] Moreover, Article 19 of the UDHR emphasizes the right of people "to seek, receive[,] and impart information."[74] These protected acts may be related to internet-related behaviors, since the right to "seek" information can be linked to browsing and searching the Internet through search engines and portals. The right to "impart" information, on the other

---

[67] *See also* Shiela Marie Rabaya, *A Pandemic of Misinformation: Legal Issues Concerning Intermediary Liability in the COVID-19 Era*, 93 (Special Online Feature) PHIL. L.J. 126, 128 (2020). The said Essay includes some parts of this discussion on freedom of expression.

[68] CONST. art. III, § 4.

[69] JOAQUIN BERNAS, THE 1987 CONSTITUTION OF THE REPUBLIC OF THE PHILIPPINES: A COMMENTARY 248 (2003 ed.).

[70] Azer Parrocha, *War vs. crime, corruption, drugs advances human rights: Palace,* PHIL. NEWS AGENCY, Dec. 10, 2018, *at* https://www.pna.gov.ph/articles/1056182.

[71] Senate of the Philippines, *De Lima: PHL cannot opt out from international obligations vs death penalty,* SENATE OF THE PHIL.: 18TH CONG. WEBSITE, Feb. 7, 2017 *at* http://legacy.senate.gov.ph/press_release/2017/0207_delima1.asp.

[72] Center for Democracy & Technology, *supra* note 66.

[73] G.A. res. 271A (III), Universal Declaration of Human Rights [hereinafter "UDHR"] (Dec. 10, 1948). "Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive[,] and impart information and ideas *through any media* and regardless of frontiers." (Emphasis supplied).

[74] *Id.*

hand, can be applied to blogging or posting information through social network sites or ISPs. Lastly, the right to "receive" information relates to the exchange of email, the reading of information through Web pages, and the downloading of information.[75] The protected rights in Article 19 are emphasized in Article 27 of the UDHR, which upholds the right of each individual to "freely participate in the cultural life of the community, to enjoy the arts[,] and to share in scientific advancement and its benefits."[76]

On top of the aforementioned UDHR provisions, Article 19 of the ICCPR provides the following protections:

> 1. Everyone shall have the right to hold opinions without interference.
> 2. Everyone shall have the right to freedom of expression; this right shall include freedom to seek, receive[,] and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice.[77]

Article 19(2) of the ICCPR mirrors the same protections in Article 19 of the UDHR. Similarly, it may be applied to expression on the Internet, given that it also includes the words "through any other media of his choice."

Like any other right, however, freedom of speech also admits of some exceptions. Both the UDHR and ICCPR provide exemptions to freedom of speech. The UDHR provides that "everyone shall be subject only to such limitations as are determined by law solely for the purpose of securing due recognition and respect for the rights and freedoms of others and of meeting the just requirements of morality, public order[,] and the general welfare in a democratic society."[78] The ICCPR, on the other hand, provides that:

> [T]he exercise of the rights [...] carries with it special duties and responsibilities. It may therefore be subject to certain restrictions, but these shall only be such as are provided by law and are necessary: (a) For respect of the rights or reputations of others; (b) For the protection of national security or of public order (ordre public), or of public health or morals.[79]

---

[75] Center for Democracy & Technology, *supra* note 66.

[76] UDHR art. 27 (1).

[77] International Covenant on Civil and Political Rights [hereinafter "ICCPR"] art. 19, Dec. 16, 1966, 999 U.N.T.S. 171.

[78] UDHR art. 29.

[79] ICCPR art. 19.

Under Philippine jurisprudence, a valid government interference to freedom of expression may be allowed if such measure passes the clear and present danger rule,[80] the dangerous tendency rule,[81] or the balancing of interests rule,[82] whichever is applicable given the circumstances.

With the continuing growth of the Internet, governments are imposing regulations on the same, threatening the medium's full potential. Some governments have passed laws prohibiting certain content on the Internet. Some even prosecute users and ISPs, or control access by blocking content directly or by insisting that ISPs impose filtering mechanisms to block illegal content.[83] The measures of control are imposed upon the intermediaries or ISPs due to easier enforcement. The rigidity of the control depends on the regime of liability adopted by the government, subject to its own constitutional policies. The adherence of these laws to freedom of speech is also dependent on the regime of liability and the particulars of the control measures adopted by the government.

In the Philippines, the fundamental right to freedom of speech cannot be whimsically subjected to regulation. Further, issues on enforceability should also be considered. Balance must be achieved.

B. Cybercrime Prevention Act and its Implementing Rules and Regulations

Intermediary liability in the Philippines can be imposed through the CPA and its IRR. Intermediaries such as Facebook and YouTube fall under the term "service provider," which is defined as "any public or private entity that provides to users of its service the ability to communicate by means of a computer system"[84] and "any other entity that processes or stores computer data on behalf of such communication service or users of such service."[85] A

---

[80] Cabansag v. Fernandez, 102 Phil. 152 (1957). The clear and present danger rule inquires as to whether words are used in such circumstance and of such nature as to create a clear and present danger that will bring about the substantive evil that the State has a right to prevent.

[81] *Id.* The dangerous tendency rule states that a person could be punished for words uttered or for ideas expressed which create a dangerous tendency, or which will cause or bring about a substantive evil which the State has a right to prevent.

[82] CARLO CRUZ & ISAGANI CRUZ, CONSTITUTIONAL LAW 522 (2015). The balancing of interests rule requires a Court to consider the circumstances in each particular case, and thereafter, it shall settle the issue of which right demands greater protection.

[83] Center for Democracy & Technology, *supra* note 66.

[84] Rep. Act No. 10175 (2012), § 3(n)(2)

[85] Rep. Act No. 10175 (2012) Rules & Regs, § 3 (ff)(2).

service provider which "willfully abets or aids in the commission of any of the offenses enumerated in [the] Act shall be held liable."[86] Liability is also imposed if service providers fail to preserve computer data within a specified period[87] or to disclose such traffic data and subscriber information after being compelled to do so by authorities.[88] The provisions on the liability of service providers seem straightforward in that the aiding and abetting must be done willfully in order to be punishable under the law.

The extent of the liability of a service provider is given more detail in the IRR of the CPA. In particular, Section 20 provides the following:

> Section 20. *Extent of Liability of a Service Provider.* – Except as otherwise provided in this Section, *no person or party shall be subject to any civil or criminal liability* in respect of a computer data for which the person or party acting as a service provider *merely provides access* if such liability is founded on:
>
> b. The obligations and liabilities of the parties under a computer data;
>
> c. The making, publication, dissemination[,] or distribution of such computer data or any statement made in such computer data, including possible infringement of any right subsisting in or in relation to such computer data: *Provided*, That:
>
> 1. The service provider does not have actual knowledge, or is not aware of the facts or circumstances from which it is apparent, that the making, publication, dissemination[,] or distribution of such material is unlawful or infringes any rights subsisting in or in relation to such material;
> 2. The service provider does not knowingly receive a financial benefit directly attributable to the unlawful or infringing activity; and
> 3. The service provider does not directly commit any infringement or other unlawful act, does not induce or cause another person or party to commit any infringement or other unlawful act, and/or does not directly benefit financially from the infringing activity or unlawful act of another person or party[.][89]

Section 20 provides an immunity from liability for service providers, subject to the three enumerated exceptions in paragraph (b). The "actual knowledge" exception in Section 20(b)(1) requires a situation wherein the

---

[86] Rep. Act No. 10175 (2012), § 5(a).

[87] § 13.

[88] § 14.

[89] Rep. Act No. 10175 (2012) Rules & Regs, § 20. (Emphasis supplied.)

service provider is aware of the unlawful facts or circumstances in relation to the material it provides access to. Section 20 is a substantial reproduction of Section 30 of the Electronic Commerce Act of 2000. If applied to the suit filed by the Bicolano tycoon, Facebook and Google Philippines may be held liable if it is proven that they were aware of the illegal nature of the videos and of the fact that they were posted on their platforms.

The CPA and its IRR adhere to the safe harbor regime of intermediary liability. An ISP will only be held liable under the IRR if they had knowledge that their platform contained illegal content. The knowledge requirement under the safe harbor regime may be actual or constructive. The IRR of the CPA requires actual instead of constructive knowledge. As will be discussed in Part IV of this Note, this choice of actual instead of constructive knowledge may pose legal and policy problems.

C. Magna Carta for Philippine Internet Freedom

The CPA has been heavily criticized since its passage. In particular, the 2014 case of *Disini v. Secretary of Justice*[90] challenged 21 sections of the CPA, with the Court ruling for the unconstitutionality of some provisions.[91] The constitutionality of the provision referring to online libel, however, was upheld as to the original author of the post but struck down with regard to those who would like or share the post .[92] The decision was met with criticism, especially from legal practitioners and members of the public advocating for the decriminalization of libel.[93] The late Senator Miriam Defensor Santiago's disagreement with the *Disini* decision regarding online libel is noteworthy[94] since this eventually led to her introduction of the "Magna Carta for Philippine Internet Freedom" ("MCPIF") bill.[95] According to her:

> [The *Disini* ruling on online libel] might precipitate libel suits related to posts on Twitter, Facebook, and Craigslist. A tweet is limited to 140 characters, and you might think that it would be difficult to commit libel with this limitation. But in a court in the United

---

[90] G.R No. 203335, 716 SCRA 237, Feb. 11, 2014.

[91] *Id.* The Court declared that Sections 4(c)(3), 12, and 19 are unconstitutional.

[92] *See* Rep. Act No. 10175 § 4(c)(4). *See also* Miriam Defensor Santiago, *Cyber Law on Libel*, 89 PHIL. L.J. 757 (2015).

[93] *See* Lorna Patajo-Kapunan, *Decriminalizing libel*, BUSINESSMIRROR, Oct. 16, 2017, *at* https://businessmirror.com.ph/2017/10/16/decriminalizing-libel./.

[94] Defensor Santiago, *supra* note 92.

[95] S. No. 53, 16th Cong., 1st Sess. (2013).

> Kingdom, the plaintiff won a libel case, because a British politician posted on Twitter.[96]

The MCPIF was written by crowdsourcing on the Internet.[97] It was filed as a response to "dangers to free speech posed by the recent Supreme Court decision upholding online libel,"[98] and to criticisms to the CPA. The main thrust of the MCPIF is to uphold the freedom of expression of Filipino citizens and to protect them justly from excessive government interference.[99] This goal is strengthened by provisions providing for court proceedings in cases where websites or networks are to be taken down, and prohibitions on censorship of content without a court order.[100]

Upon perusal of the bill, however, it can be seen that there is no specific provision referring to intermediary liability of ISPs. Despite the absence of this provision in the bill, Senator Defensor Santiago has acknowledged the necessity of a safe harbor provision for intermediary liability:

> Because of the dangers to free speech posed by the recent Supreme Court decision upholding online libel, I have filed a new bill in the Senate entitled "Magna Carta of Internet Freedom," which was written by crowdsourcing in the Internet. In light of the recent Supreme Court decision, *I highly recommend that the Congress protect online service providers from liability for the posts made by their users.* This is called the "Safe Harbor Provision," under the U.S. Communications Decency Act.[101]

Senator Defensor Santiago made reference to Section 230 of the CDA as an example of legislation protecting online service providers. However, she also recognized the abuse that may be perpetuated due to the same provision. In particular, Senator Defensor Santiago noted that "[w]e have to, however, restudy the 'Safe Harbor Provision' [of Section 230]

---

[96] Defensor Santiago, *supra* note 92, at 759. *See Twitter libel Caerphilly councillor pays rival £3,000*, BBC NEWS, Mar. 10, 2011, *at* http://www.bbc.com/news/uk-wales-south-east-wales-12704955; *Twitter claim costs councillor £53,000*, WALES ONLINE, Mar. 10, 2011, *at* https://www.walesonline.co.uk/news/wales-news/twitter-claim-costs-councillor-53000-1845308.

[97] Defensor Santiago, *supra* note 92, at 760.

[98] *Id.*

[99] Senate of the Philippines, *Magna Carta for Internet Freedom to Replace Anti-Cybercrime Law*, SENATE OF THE PHIL.: 18TH CONG. WEBSITE, Nov. 20, 2012, *available at* http://www.senate.gov.ph/press_release/2012/1130_santiago1.asp.

[100] *Id.*

[101] Defensor Santiago, *supra* note 92, at 760. (Emphasis supplied.)

because it can be abused. The provision exempts the website from liability, while its operators shield posters by means of coding that allows people to post anonymously."[102] Hence, the adoption in the Philippines of a provision similar to Section 230 may lead to courts using the provision to dismiss complaints for invasion of privacy, misappropriation of trade secrets, cyberstalking, and negligence.

The introduction of the MCPIF and Senator Defensor Santiago's reference to Section 230 of the CDA happened before the passage of the IRR of the CPA. It is in this IRR where ISPs who merely give access to third party content are given immunity from liability, save for several exceptions. As proposed in this Note, said immunity provided in the IRR is not a true safe harbor since it disincentivizes regulation by ISPs. Likewise, adopting a provision similar to Section 230 may lead to the problems pointed out by Senator Defensor Santiago.

The passage of a true safe harbor provision that balances freedom of expression with the need to regulate ISPs is consistent with the thrust of the MCPIF. The MCPIF, however, was not signed into law. On March 16, 2015, the bill was consolidated and submitted in Committee Report No. 113,[103] which later became Republic Act No. 10844.[104] This law is otherwise known as the "Department of Information and Communications Technology Act of 2015," which adopts the provisions of MCPIF constituting a similar department. The other provisions of the MCPIF, however, were not signed into law.[105]

IV. ANALYSIS

Criticisms on the CPA primarily center on its possible suppression of freedom of speech and expression. The most controversial provision is Section 4(c)(4) of the CPA, which applies criminal libel to acts "committed through a computer system or any other similar means which may be devised

---

[102] *Id.*

[103] C. Rpt. 113, 16th Cong., 3rd Sess. (2016). Committees on Science and Technology; Civil Service, Government Reorganization, and Professional Regulation; Constitutional Amendments and Revision of Codes; Finance, *available at* http://legacy.senate.gov.ph/lis/committee_rpt.aspx?congress=16&q=113.

[104] Rep. Act No. 10844, § 1. Department of Information and Communications Technology Act of 2015 (2016).

[105] *See* S. No. 53, 16th Cong., 1st Sess. (2013). *Magna Carta for Philippine Internet Freedom, available at* http://legacy.senate.gov.ph/lisdata/1586313101!.pdf.

in the future."[106] Traditional libel in the RPC has already been met with criticisms over the years, with some claiming that maintaining libel as a criminal offense runs afoul not only of the Constitution, but also of the Philippines' international obligations.[107] The public outcry has since then extended to online libel through the CPA.[108]

The problematic provisions for intermediary or ISP liability provided in the CPA and its IRR, however, are often overlooked. Section 20 of the IRR providing for the general immunity of ISPs from liability for content that it "merely provides access" to is substantially reproduced from Section 30 of the Electronic Commerce Act. Likewise, the exemptions provided—with particular emphasis on the actual knowledge exception—are reproduced from the same law. Since 2000, provisions for intermediary liability have then been present in the legal landscape of the Philippines. However, these provisions have not been tested prior to the case filed against Facebook and Google Philippines.[109] The following discussion will illustrate how a plain-text reading of the current form of the IRR of the CPA can lead to the problems realized by the U.S. during the case of *Stratton*.[110] Possible legal remedies to the disincentivizing effect of current Philippine laws on intermediary liability will also be explored.

A. The "Safe Harbor" of the Cybercrime Prevention Act

The provisions on intermediary liability in the CPA adhere to the safe harbor regime, wherein intermediaries are only held liable for defamatory or illegal content if they had knowledge that their platform contained illegal content.[111] "Knowledge" under the safe harbor regime may be actual or constructive, and Section 20 of the IRR of the CPA requires actual knowledge for liability to attach. As a result, most ISPs in the Philippines may opt to not impose its own screening measures over content posted on its platform. Instead, ISPs impose "notice and takedown" measures to allow its users to report illegal or defamatory content. The responsibility to screen illegal or defamatory content is then passed to the users.

---

[106] Rep. Act No. 10175 (2012), § 4(c)(4).

[107] Kelvin Lester Lee & Juan Paolo Villonco, *An Examination of Cyberlibel in the Philippines: A Study of the Current State of Online Defamation*, 57 ATENEO L.J. 1084 (2013).

[108] Human Rights Watch, *Philippines: New 'Cybercrime' Law Will Harm Free Speech*, Sept. 28, 2012, *at* https://www.hrw.org/news/2012/09/28/philippines-new-cybercrime-law-will-harm-free-speech.

[109] Punay, *supra* note 13.

[110] *Stratton Oakmont*, 23 Media L. Rep. (BNA) 1794, 1796-98.

[111] Fernandez & Pangalangan, *supra* note 38, at 772, *citing* Larusdottir, *infra* note 113, at 476.

For instance, YouTube and Facebook contain the following notice and takedown measures:

FIGURE 1. Example of YouTube's Notice and Takedown Measure

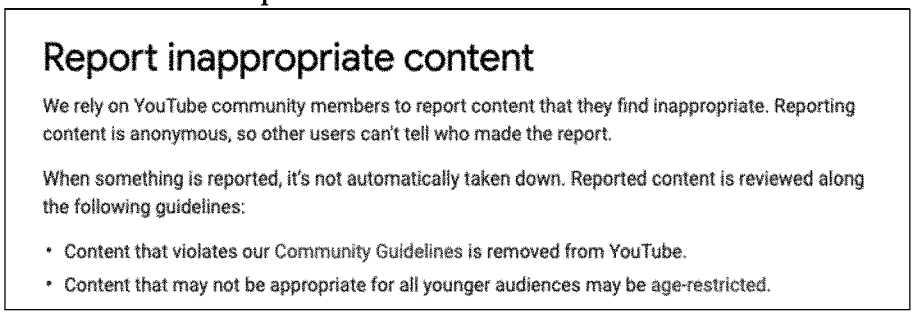
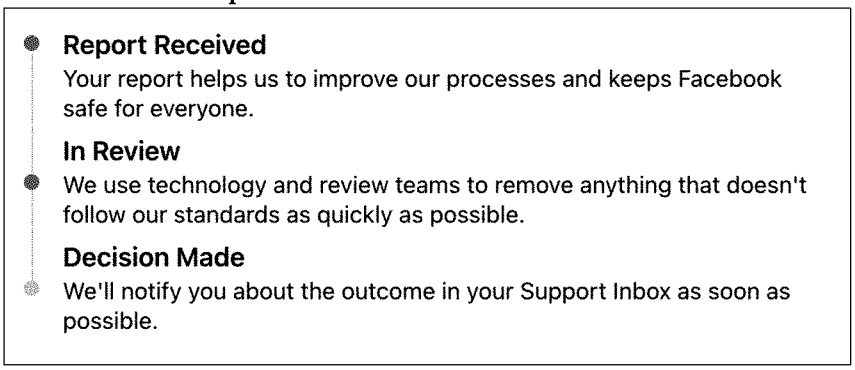


FIGURE 2. Example of Facebook's Notice and Takedown Measure



The safe harbor regime adopted in the Philippines poses the same problems faced in the U.S. jurisdiction prior to Section 230 of the CDA. Before the ruling in *Stratton* and the subsequent passage of Section 230 of the CDA, courts refused to impose liability on websites that knew nothing of its content. The ruling in *Stratton*, however, showed that courts were inclined to punish "Good Samaritan" ISPs who imposed regulatory measures but are not able to filter all illegal content, rather than ISPs that did not impose regulatory measures at all. With this reasoning, it is then better for ISPs to not regulate its content at all than to impose any screening measures, or what has been referred to as the *Stratton* dilemma. Failed attempts at screening content are punished far more harshly than the absence of any kind of screening measure. The disincentivizing nature of the legal regime prior to Section 230 is then apparent. Similarly, Section 20 of the IRR of the CPA provides general immunity to a service provider which "merely provides access" to data. There

are three exceptions provided to this general immunity, and the one which may lead to the *Stratton* dilemma is the actual knowledge exception.[112]

In its adherence to the safe harbor regime, the IRR of the CPA could have adopted either actual or constructive knowledge of illegal or defamatory content as an exception to the general immunity provided by the law. The actual knowledge exception was the one eventually written in the IRR, although this might not have been the most effective choice. It is argued by some authors that adopting the constructive knowledge requirement in the safe harbor regime will actually eliminate the *Stratton* dilemma.

Scholars arguing for the constructive knowledge approach point out that "[i]mposing the actual knowledge standard would lead to a low risk of liability for the [intermediaries], as in that case it must be established that the [intermediary] actually knew about the infringing material in order to trigger the potential liability."[113] "Actual" knowledge is dependent on users who utilize the notice and takedown mechanism of the ISP. The actual knowledge exception perpetuates the *Stratton* dilemma since it provides ISPs with an incentive to not monitor the content hosted in its facilities. Applying the constructive knowledge approach, however, negates this problem since it imposes a higher risk of liability for intermediaries. Actual knowledge of the illegal content need not be proven. Hence, the constructive knowledge approach forces intermediaries to enact mechanisms to avoid liability,[114] without adhering to the even more problematic strict liability regime. The strict liability regime cannot be adopted in the Philippines since this may produce a chilling effect that is against the freedom of expression and speech protected by the Constitution.[115]

Applying the constructive knowledge approach, however, may pose some difficulties. First, in order to maximize the effectiveness of the constructive knowledge approach, standards on determining the existence of "constructive" knowledge must be imposed. Also, a potential problem with the constructive knowledge approach is that it may lead to mass take-downs

---

[112] Rep. Act No. 10175 (2012) Rules & Regs, § 20(b)(1). The actual knowledge exception states the following: "The service provider does not have actual knowledge, or is not aware of the facts or circumstances from which it is apparent, that the making, publication, dissemination[,] or distribution of such material is unlawful or infringes any rights subsisting in or in relation to such material[.]"

[113] Jonina Larusdottir, *Liability of Intermediaries for Copyright Infringement in the Case of Hosting on the Internet,* 47 SC. ST. L. 471, 477 (2004).

[114] *Id.*

[115] Fernandez & Pangalangan, *supra* note 38.

by ISPs, since knowledge and liability can easily be imposed upon them. The more stringent standard of potentially imposing liability due to constructive knowledge may also lead to the chilling effect associated with a strict liability regime.

B. The "General Immunity" of Section 230 of the Communications Decency Act

As a response to the *Stratton* dilemma, the U.S. Congress enacted Section 230 of the CDA, thereby expressly relieving ISPs from liability as publishers in two ways. First, it immunizes ISPs from the *Stratton* dilemma, which means that an ISP cannot be held liable as a publisher just because of its attempt to filter its content.[116] Hence, it presumably removes the disincentive to regulation, which is in line with the law's underlying purpose "to remove disincentives for the development and utilization of blocking and filtering technologies[.]"[117] Second, an ISP is immune from liability to those whose content it filters, although the content is constitutionally protected. In sum, websites would not face liability for not blocking enough content or for blocking too much content.[118]

While the intent of Section 230 of the CDA is to eliminate the *Stratton* dilemma and to allow the unimpeded growth of the Internet,[119] the provision faced many criticisms due to the apparent blanket immunity that it provided to ISPs. The main point of criticism was centered on the apparent "do-nothing" approach fostered among ISPs.[120] With this approach, the disincentive seen in *Stratton* may be eliminated, but there remains no *incentive* to actual self-regulation by ISPs. It is true that they may not be held liable for their attempts at self-regulation, however, they will also not be held liable for knowingly hosting illegal or defamatory content. ISPs will enjoy the protection of Section 230, whether or not they make use of any mechanism for self-regulation.[121] This was the interpretation of Section 230 promulgated by the Court in *Zeran*, since the said ruling extends Section 230's immunity to distributor liability.[122] Even if the ISP only engages in distributor activities, without contributing to the illegal content *and* without implementing any kind of screening measure, it is still exempt from liability.

---

[116] French, *supra* note 31, at 450.
[117] Sevanian, *supra* note 52, at 136, *citing* 7 U.S.C. § 230(b)(4).
[118] French, *supra* note 31, at 450.
[119] 141 Cong. Rec. H8468 (daily ed. Aug. 4, 1995) (statement of Rep. Cox).
[120] Sevanian, *supra* note 52, at 136.
[121] *Id.* at 136.
[122] *Zeran*," 129 F.3d 327 (4th Cir. 1997).

Zeran is the case where Section 230 was first interpreted by a circuit court. In this case, Kenneth Zeran became the victim of an Internet hoax after an unknown perpetrator claimed to be him on an online AOL bulletin board. On this forum, the perpetrator began advertising shirts and other merchandise supporting and glorifying the bombing of the Oklahoma City federal building in 1995. Zeran received dozens of threatening phone calls due to the posts. He immediately demanded that AOL take down the posts. However, even after being notified, AOL failed to prevent the continued postings of the perpetrator. Hence, the subsequent suit filed by Zeran.[123]

In response to the suit filed by Zeran, AOL claimed immunity under Section 230 of the CDA. Zeran, on the other hand, contended that Section 230 only applied to the very precise notion of "publisher" and that the liability of AOL is predicated on its role as "distributor." Zeran then postulated that "he was not trying to place AOL in the role of publisher, and therefore it did not qualify for [Section] 230's protections."[124] The Fourth Circuit, however, ruled in favor of AOL. It found that the distributor classification is merely a subset of the broad common law notion of publisher. It concluded that distributor liability also threatened the congressional goal of preventing websites from avoiding attempts to screen content. In effect, the court extended the immunity given by Section 230 to distributor liability. This became known as the "third party immunity" interpretation, where ISPs are immune from liability for hosting *any* kind of content from third persons.

It is apparent, however, that the plain language of Section 230 negates the applicability of the provision to distributor liability.[125] The word "publisher" is in the law but this was followed by "speaker" and not distributor. Immunity from distributor liability is not expressly provided. A blanket immunity from Section 230 conflicts with the Congress' aim of encouraging self-regulation by ISPs because an incentive *against* self-regulation would replace the disincentive realized in *Stratton*. The *Zeran* approach will encourage ISPs to let go of any form of self-regulation as this can cut operational costs in the long run.[126] Moreover, the *Zeran* ruling has the opposite effect of what Congress intended, since it encourages ISPs to do nothing. Taking steps to regulate its content provides no additional immunity.[127] It is then unfortunate that, although some subsequent decisions

---

[123] French, *supra* note 31, at 452.
[124] *Id.* at 453.
[125] Sevanian, *supra* note 52, at 136.
[126] *Id.* at 137.
[127] French, *supra* note 31, at 466.

attempted to break free from the *Zeran* interpretation,[128] U.S. jurisprudence has mostly followed the *Zeran* decision.[129]

C. The True "Safe Harbor": The Remedy to the Disincentivizing Intermediary Laws

The "safe harbor" of Section 20 of the IRR of CPA disincentivizes regulation, much like the pre-Section 230 era in the U.S. jurisdiction. It promotes a hands-off approach among ISPs, since self-regulation may lead to the "actual knowledge" exception to general immunity. Adopting a "general immunity" approach similar to Section 230 of the CDA, however, does not precisely solve the problem posed in *Stratton.* While Section 230 invariably promotes the constitutionally enshrined freedom of speech, the *Zeran* interpretation adopted by most courts perpetuates rather than solves the disincentivizing effect present in intermediary liability legislation. The intention of the legislators to promote the unimpeded growth of the Internet just opened the floodgates for abuse of the online platform.[130] Lastly, strict liability is not an option since it would definitely produce the chilling effect that the Bill of Rights has long-protected freedom of expression from.[131] It can then be concluded that a middle-ground between the problematic immunity accorded by the *Zeran* interpretation of Section 230 of the CDA and the *Stratton*-like disincentive present in Section 20 of the IRR of the CPA is necessary.

With the foregoing conclusions, there is an apparent deadlock as to the proper approach to intermediary liability. A plain-text reading of Section 20 of the IRR of the CPA leads to the *Stratton* dilemma, although an interpretation by Philippine courts may solve the problem. Several approaches have already been explored by scholars as a response to the problems posed by Section 230 of the CDA. These approaches can likewise apply to the Philippine legal landscape. The first is an approach requiring constructive instead of actual knowledge in imposing liability to ISPs.[132] The merits of this approach have been discussed at the start of Part IV, although such an approach may also pose problems similar to a chilling effect.

---

[128] *Id.*

[129] *Id.*

[130] Tischler, *supra* note 53, at 277-78. Several letters were sent by Congressional leaders proposing an amendment to Section 230. "Citing interference with their ability to investigate and prosecute child prostitution and sex trafficking, the Attorneys General suggested amending the statute to eliminate protection from state criminal statutes."

[131] Fernandez & Pangalangan, *supra* note 38, at 772.

[132] Larusdottir, *supra* note 113.

Moreover, there may be difficulty in using the constructive knowledge approach in the Philippines given the use of "actual knowledge" in the text of the IRR of the CPA. While it may be argued that the phrase "or is not aware of the facts or circumstances from which it is apparent"[133] following the words "actual knowledge" in Section 20 gives way to an interpretation accepting constructive knowledge as such "awareness," this is still dependent on a definite ruling by Philippine courts favoring this interpretation.

It is thus imperative to explore other approaches that may steer the Philippine jurisdiction towards a true "safe harbor" that can balance the people's fundamental right to speech with the increasing need to regulate Internet content. Looking at the suggested remedies to combat the criticisms to the apparent general immunity provided by Section 230 of the CDA may help in developing remedies to the disincentive to self-regulation present in Philippine law. These remedies are interspersed throughout rulings of several U.S. District Courts. Mostly, the exceptions take the form of state legislation in the U.S.:

> In all, as a matter of upholding Congress' intent to encourage ICS self-regulation, a synthesis of the aforementioned sources reveals that an ICS could lose its Section 230 "good Samaritan" immunity status by either (1) engaging in "bad faith" by "encourag[ing]" or "solicit[ing]," or partaking in the "creat[ion] or develop[ment]," [of] illegal or offensive third party content; (2) "willingly" implementing a system in which the ICS does not screen for the identity of third party users who post illegal or offensive content on its website; (3) promising to remove illegal or offensive content from its website, but then failing to do so; or (4) failing to engage in self-regulation, as required by Section 230-consistent state laws.[134]

The remedies are in the form of exceptions to the general immunity of Section 230. These exceptions focus on the existence of bad faith in hosting illegal content, and in partaking in the creation of illegal content. An analysis of these exceptions will show that the first two are "publisher" activities, in line with the distributor-publisher distinction prior to Section 230. An ISP engaging in publisher activities *that are not screening measures* and contributing to illegal content would not then be entitled to immunity from intermediary liability. Hence, Section 230 immunity would be limited to ISPs exercising publisher activities in good faith and in relation to self-regulation. This is also in line with the fact that Section 230 immunity only applies to "interactive

---

[133] Rep. Act No. 10175 (2012) Rules & Regs, § 20(b)(1).

[134] Sevanian, *supra* note 52, at 139. (Citations omitted.)

computer services" and not to "information content providers," and that the immunity in Section 20 of the IRR of the CPA only extends to service providers that "merely provides access" to illegal content. In fine, an ISP acting as a mere distributor cannot be treated as a publisher or speaker; however, an ISP engaging in publisher activities contributing to illegal content cannot be given the same immunity. Other approaches are definitional immunity, promissory estoppel, right of reply, and the totality of circumstances approach. Number (4) in the quoted exceptions also points to possibly conditioning the immunity provided by Section 230 to the existence of self-regulating measures by ISPs.

1. Definitional Immunity

Under the definitional immunity approach, Section 230 can be read as a definitional clause rather than as an immunity from liability. With this interpretation, "[Section] 230 offers no immunity unless an entity qualifies as an interactive service provider that did not contribute content, *and* has taken Good Samaritan actions"[135] to regulate its content. The immunity would only be given to ISPs that do not contribute illegal content, and who actually take measures to regulate illegal content posted in its platform. This approach does not recognize any protection, without Good Samaritan efforts, to screen content. The disincentives to regulation present prior and even *after* the passage of Section 230 are then effectively eradicated. Section 230 would fully serve its purpose as a Good Samaritan Law.

Applied in the Philippines, the wording of the CPA and Section 20 of the IRR may be interpreted as a definitional clause. Section 20 extends the immunity to ISPs which "merely provides access" to illegal or defamatory content. Hence, the wording is already parallel to an interactive service provider that did not contribute content in Section 230. An interpretation by Philippine courts or a revised IRR may add the "Good Samaritan" requirement. However, given that the Philippines is a civil rather than a common law country, there may be difficulties in stretching the definition beyond what can be readily gleaned from the wording of the law. The definitional immunity approach may thus be difficult to apply in Philippine jurisdiction, absent a court decision or an amendment of the law or IRR.[136]

---

[135] French, *supra* note 31, at 465–66.

[136] CIVIL CODE, art. 8. "Judicial decisions applying or interpreting the laws or the Constitution shall form part of the legal system of the Philippines."

2. Promissory Estoppel

The promissory estoppel exception to Section 230's general immunity relies upon contract principles. This was first tested in the case of *Barnes v. Yahoo!, Inc.*,[137] wherein Yahoo's Director of Communications assured the plaintiff that Yahoo would "take care" of the indecent and defamatory profile posted by the plaintiff's ex-boyfriend. Yahoo, however, failed to take down the said defamatory content. In this case, the circuit court indeed found that Yahoo was immune under Section 230. However, the court also ruled that the immunity did not preclude the plaintiff from suing Yahoo based on a state law contract claim of promissory estoppel. In fine, the plaintiff can sue Yahoo due to the latter's failed promise of removing the indecent profile, and Barnes' reliance on that promise.[138]

Prior to the *Stratton*[139] decision and the subsequent passage of Section 230 of the CDA, the principle of promissory estoppel would be difficult to apply. In this pre-Section 230 situation, an ISP's refusal to remove or edit third-party content is understandable, given that there was no immunity for Good Samaritan efforts. "[I]f the ISP did remove or edit certain content, it could be held liable for other content that it did not remove."[140] It was only under the CDA when the ISPs were expressly protected from this kind of liability. The case of *Barnes*, which was decided after the passage of Section 230, clarified the applicability of the principle of promissory estoppel. "In light of *Barnes*, ISPs will likely be more careful than ever when addressing requests to remove defamatory third-party content."[141]

The principle of promissory estoppel is also present in the Philippine jurisdiction. Under this doctrine:

> [E]stoppel may arise from the making of a promise, even though without consideration, if it was intended that the promise should be relied upon and in fact was relied upon, and if a refusal to enforce it would be virtually to sanction the perpetration of fraud or would result in other injustice.[142]

---

[137] 570 F. 3d 1096, 1109 (9th Cir. 2009).

[138] *Id.*

[139] 23 Media L. Rep. (BNA) 1794, 1796-98 (N.Y. Sup. Ct. 1995).

[140] Ali Grace Zieglowsky, *Immoral Immunity: Using a Totality of the Circumstances Approach To Narrow the Scope of Section 230 of the Communications Decency Act*, 61 HASTINGS L.J. 1307, 1329 (2009).

[141] *Id.* at 1330.

[142] Ramos v. CBP, G.R. No. 29352, 41 SCRA 565, 588, Oct. 4, 1971.

Promissory estoppel is a civil law concept that can then apply against ISPs, on top of the current legislation on intermediary liability.

3. Right of Reply and Counter-Notice Mechanism

The right of reply is another remedy suggested to counter the problematic aspects of Section 230 of the CDA. Specifically, it is proposed that the immunity under Section 230 must be conditioned on providing a right of reply to third persons who are the subject of a defamatory content in an ISP's platform.[143] Such proposed amendment to Section 230 is in line with the "safe harbor" provisions of the Digital Millennium Copyright Act ("DMCA").[144]

It may be argued that a regime of intermediary liability with a "right of reply" or a "notice and takedown" mechanism provides the true safe harbor for ISPs. Unlike the purported safe harbor of the current Philippine laws on intermediary liability, actually providing either a "right of reply" or a "notice and takedown" mechanism may eradicate the disincentivizing effect of current laws. With the right of reply in place, ISPs would be obligated to provide space for replies of third persons who are subjects of defamatory content in the platform. The original statement will remain accessible to Internet users who may then consider both the original, allegedly defamatory content, and the reply.[145] Further, a right of reply may be more effective than a notice and takedown mechanism[146] since the latter may suppress freedom of speech due to the discretion given to ISPs in taking down third-party content. In effect, notice and takedown mechanisms may lead to third parties losing their platform for expressing their thoughts. On the contrary, the availability of both the original, allegedly defamatory content and the reply would provide a platform for both sides. Internet users would then be given the opportunity to make up their own minds as to the truthfulness of the competing statements.[147]

Further, the DMCA also provides a counter-notice mechanism. Once the content reported through the notice and takedown mechanism is removed, "the website operator must notify the poster of the alleged infringing material and allow the poster the opportunity to file a counter-

---

[143] Michael D. Scott, *Would a Right of Reply Fix Section 230 of the Communications Decency Act*, 4 J. INT'L MEDIA & ENT. L. 57, 64–65 (2011).

[144] *Id. See* Digital Millennium Copyright Act, Pub. L. No. 105-304, 112 Stat. 2860 (1998), codified at 17 U.S.C., § 512 (2010).

[145] Scott, *supra* note 143, at 67–68.

[146] *Id.* at 67.

[147] *Id.* at 68.

notice to dispute the removal."[148] This mechanism, on top of the right of reply, would also give both the subject of the alleged illegal content and the poster an opportunity to be heard. However, such a mechanism may also pose problems in the form of a "'heckler's veto,' whereby people complain about speech because they dislike the speakers or object to [the] views [thereof]."[149] Any "heckler" may request the removal of third-party content, which could also suppress protected speech.[150] The absence of an appeals process after the removal of content has been one of the most persisting criticisms against Facebook in recent years.[151] Facebook responded to these criticisms by passing its Community Standards in 2018,[152] which contain a provision for an appeals process.

As applied in the Philippines, an amendment in the CPA and its IRR providing for a mandatory right of reply or a notice and takedown mechanism may provide a true safe harbor regime for intermediaries. This is a legislative solution which passes the constitutional standards of freedom of speech, while still allowing a modicum of regulation over intermediaries. Currently, most intermediaries in the Philippines already impose notice and takedown mechanisms.[153] However, the missing factor to complete the DMCA model is actually conditioning the immunity provided by law with the fulfillment of these mechanisms. This model is similar to the definitional immunity approach since immunity from liability will be conditioned upon the fulfillment of Good Samaritan requirements. Such requirements, however, are

---

[148] Bolson, *supra* note 47, at 14. *See* 17 U.S.C. § 512(g)(2)(A); 17 U.S.C., § 512(g)(2)(B).

[149] Bolson, *supra* note 47, at 15. (Citations omitted.) *See* DANIELLE KEATS CITRON, HATE CRIMES IN CYBERSPACE 179 (2014).

[150] *Id.*

[151] Washington Post, *Facebook reveals its censorship guidelines for the first time – 27 pages of them*, L.A. TIMES, Apr., 24, 2018, *at* https://www.latimes.com/business/ technology/la-fi-tn-facebook-guidelines-20180424-story.html.

[152] Facebook Community Standards, *at* https://www.facebook.com/ communitystandards/ (last accessed May 20, 2020). *See* Todd Haselton, *Here's Facebook's once-secret list of content that can get you banned*, CNBC, Apr. 24, 2018, *at* https://www.cnbc.com/ 2018/04/24/facebook-content-that-gets-you-banned-according-to-community-standards.html; Google Legal Help, Report Content for Legal Reasons, *at* https://support.google.com/ legal/answer/3110420?hl=en (last accessed May 20, 2020); YouTube Help, Submit a copyright takedown request, *at* https://support.google.com/youtube/answer/2807622?hl=en (last accessed May 20, 2020); Twitter Archive Eraser, *Twitter: your account has been locked due to DMCA takedown notice! Here is how to cleanup and unblock your account*, June 28, 2020, *at* https://twitterarchiveeraser.medium.com/delete-tweets-dmca-b17e0181c7; Instagram, *How does Instagram process United States Digital Millennium Copyright Act (DMCA_ counter-notifications?*, *at* https://www.facebook.com/help/instagram/697328657009330?helpref=related (last accessed May 20, 2020).

[153] Google Legal Help, *supra* note 152; YouTube Help, *supra* note 152.

specific to the right of reply and counter-notice procedures. The disincentivizing effect of the current law would then be eradicated due to these additional requirements.

4. The Manila Principles

The aforementioned remedies eventually found its way in a set of six principles created by civil society organizations from around the world. These principles became known as the Manila Principles, which focus on best practices for intermediaries rather on a certain set of rules.[154] The six principles are as follows:

1. Intermediaries should be shielded from liability for third-party content.
2. Content must not be required to be restricted without an order by a judicial authority.
3. Requests for restrictions of content must be clear, be unambiguous, and follow due process.
4. Laws and content restriction orders and practices must comply with the tests of necessity and proportionality.
5. Laws and content restriction policies and practices must respect due process.
6. Transparency and accountability must be built into laws and content restriction policies and practices.[155]

While the Manila Principles are not binding in the same manner as statutes are, they have become influential in the reform of intermediary laws in several parts of the world. These Principles have been identified as "essential principles that should guide any intermediary liability framework."[156] For example, the Manila Principles have been used as the basis for recommended amendments to the DMCA notice and takedown regime. The focal point in the suggested reform and a consistent recommendation among scholars forwarding the Manila Principles is the idea that content should only be restricted via court order.[157] This idea puts forward the notion that intermediaries should not be made arbiters of what is illegal or

---

[154] Emily Laidlaw & Hilary Young, *Internet Intermediary Liability in Defamation*, 56 OSGOODE HALL L.J. 112, 138 (2018).

[155] Electronic Frontier Foundation, *Manila Principles on Intermediary Liability*, *at* www.manilaprinciples.org/principles (last accessed May 20, 2020).

[156] Laidlaw & Young, *supra* note 154, at 139, *citing* David Kaye (Special Rapporteur), *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression,* UNHRC, 38th Sess., ¶ 14, UN Doc A/HRC/38/35 (2018).

[157] *Id.*

defamatory. The determination of whether or not an act is illegal is within the realm of law, which is clearly beyond the sphere of intermediaries.[158]

The idea forwarded by the Manila Principles, however, is not infallible. The slow and tedious court processes in the Philippines, in particular, may be an insurmountable challenge in implementing the rule that content takedowns should only be via court order. A compromise may then be essential, such as the formation of an independent body focused on the determination of the illegality of third-party content. Further, an appeals process in line with the counter-notice mechanism forwarded by the DCMA may actually remain to be the most expedient regulatory practice. The court order requirement of the Manila Principles or the alternative independent body may then be better utilized in this appeals process.

V. CONCLUSION AND RECOMMENDATIONS

It is evident that balancing the people's freedom of speech and expression with the need to regulate third party content in ISPs is a difficult task. Many solutions on how Section 230 can be improved have been forwarded, as elaborated in the previous part of this Note. There is also an increasing clamor to do away with Section 230 altogether, given that the internet startups it sought to protect have become giants.[159] The initial purpose of the law is then moot and the provision may now do more harm than good. The same sentiments forwarded for the amendment or repeal of Section 230 is applicable in Philippine jurisdiction, given that the provisions of the CPA and its IRR on intermediary liability pose the same problems as that of the pre-Section 230 era. In particular, the disincentives for regulation similar to the *Stratton* dilemma are present due to Section 20 of the IRR of the CPA. Hence, adopting a legal regime similar to Section 230 of the CDA would not be the proper remedy to the current regime of intermediary liability in the Philippines that is problematic.

Out of all the legal remedies discussed in this Note, a legal regime similar to the DMCA and the ideas forwarded by the Manila Principles seem to be the mechanism that can achieve the balance between freedom of expression and necessary regulation. To contextualize such an approach, it is useful to examine the content regulation measures employed by Facebook.

---

[158] *Id.*

[159] Issie Lapowsky & Steven Levy, *Here's What Facebook Won't Let You Post*, WIRED, Apr. 24, 2018, *at* https://www.wired.com/story/heres-what-facebook-wont-let-you-post.

On top of notice and takedown measures, Facebook and Google are also known to employ "content moderators" who comb the "vastness of cyberspace in a virtual search-and-destroy mission to expunge deeply disturbing images and videos.[160] These content scavengers examine the materials uploaded daily, and delete disturbing material following the community guidelines of U.S. social media platforms.[161] Users can also flag inappropriate content, and the content moderators will assess[162] whether such flagged content violated community standards.[163] It is then apparent that Facebook does not rely entirely on flags or notice by third persons.

Moreover, Facebook also allows users to appeal bans on individual posts and entire Pages. Facebook promises a speedy clarification and a possible reconsideration,[164] which are in line with the counter notice requirement of the DMCA. This mechanism then recognizes the freedom of speech and expression of the third-party content provider, subject to limitations enunciated in the Community Standards. The publication of the Community Standards is in line with the fifth and sixth Manila Principles. These Principles focus on transparency and accountability, with the fifth Principle stating that "[l]aws and content restriction policies and practices must respect due process."[165] The sixth Principle, on the other hand, states that "[t]ransparency and accountability must be built into laws and content restriction policies and practices."[166] Content restriction policies, such as the Community Standards of Facebook, must then be made public so that content providers and users alike are aware of what is acceptable in the platform.

Legally, Facebook is under no obligation to write policies regulating content posted in its platform. It is protected by the immunity provided by Section 230 of the CDA. However, the passage of its 27-page Community Standards[167] is an example of how an ISP can go beyond the minimum requirements imposed by law. Facebook chose to "keep itself from descending into a snake pit of harassment, bullying, sexual content[,] and gun-running."[168] The standards were also published as a response to the constant

---

[160] Mariejo Ramos, *'Cyber cleaners' in PH: A dirty job, but someone's got to do it*, INQUIRER.NET, Nov. 18, 2018, *at* https://technology.inquirer.net/81316/cyber-cleaners-in-ph-a-dirty-job-but-someones-got-to-do-it.

[161] *Id.*

[162] *See* FIGURE 2, *supra* p. 161, which makes reference to such acts by content moderators.

[163] Facebook Community Standards, *supra* note 152.

[164] *Id.*

[165] Laidlaw & Young, *supra* note 154, at 138.

[166] *Id.*

[167] Facebook Community Standards, *supra* note 152.

[168] Lapowsky & Levy, *supra* note 159.

criticism directed at Facebook due to its seemingly lax control over the content posted on its platform.

Not all ISPs, however, are willing to go beyond the minimum regulatory mechanisms required by law. Facebook and Google are ISP giants who can afford these additional screening measures. ISPs vary in terms of reach to users and technical capabilities to moderate content posted on their platform. An amendment of the CPA and its IRR, or the repeal of the CPA in favor of a legislation similar to the MCPIF may bring the true "safe harbor" for users and ISPs in the Philippines. A provision similar to the DMCA model for intermediaries containing notice and takedown mechanisms, a right of reply, and a counter-notice or appeals procedure will safeguard the freedom of expression of third-party content providers and protect the subjects of defamatory or illegal content. Further, ISPs must be required to publish their content restriction policies to accord transparency and accountability in their restriction of third-party content. These restrictions, after all, are heavily entwined with the constitutionally protected freedom of speech and expression. Lastly, courts may also utilize the contract principle of promissory estoppel, regardless of the legal regime on intermediary liability adopted by law.

Constitutional and contract principles may then be utilized to safeguard the people's freedom of expression and provide a true "safe harbor" for ISPs who exert Good Samaritan efforts to regulate content posted in its platform. A DMCA model following the definitional immunity approach will ensure that the rights of all key parties—namely the third-party content provider, the subject entity of the content posted, and the ISP—are protected. Above all considerations, the freedom of expression enshrined in the Bill of Rights will be protected. The pen will then remain mighty, and its power will be properly tempered by law.

- o0o -