LET THE USERS BE THE FILTER? CROWDSOURCED FILTERING TO AVOID ONLINE INTERMEDIARY LIABILITY

Ivar A. Hartmann[*]

I. Introduction

Online platforms for decentralized content production or for plain social interaction constitute one of the fundamental frontiers of innovation on the internet. Companies and other entities contribute to this by designing the system and maintaining it in their servers, while also takings steps to guarantee that internet users can make the best out of such environments. That is to say, although the purpose of such companies is to profit from user-generated content or to simply let individuals co-exist in communion, they play a crucial role as intermediaries. Because they are the managers of online communities where – just like in the real world – law infringement can occur, these companies are constantly sued by users themselves or third parties under the allegation that they have a responsibility for what is done in their platform.

Even though safe harbor provisions exist in American and European Law that release intermediaries from a duty to proactively monitor and filter user activity on their platform, the liability standard is constantly shifting. Copyright owners pleas, for example, demanding a different, less passive role for intermediaries has been gaining ground recently. The most poignant examples of that in recent times are the U.S. Court of Appeals for the Second Circuit's decision in April 2012 that overturned a summary judgment dismissing Viacom's case against YouTube[1] and the "right to be forgotten" ruling by the Court of Justice of the European Union[2].

This creates an environment where safe harbor provisions no longer offer enough protection against the high risk of liability. Engaging in full-fledged filtering, on the other hand, has its problems. As a result, intermediaries find themselves between a rock and a hard place.

This article describes such setting – where intermediaries have incentives both to filter and not to filter content on their platforms – and outlines a few arguments why enabling and encouraging users themselves to filter content on platforms could present itself as a solution to intermediaries' problems. It is not intended as an exhaustive enumeration of the arguments in favor and against having users themselves filter content

---

[*] Professor at FGV Law School (Rio de Janeiro, Brazil). S.J.D. candidate, Rio de Janeiro State University (Brazil). LL.M., Harvard Law School. MSc, Pontifical Catholic University of Rio Grande do Sul (Porto Alegre, Brazil). I would like to especially thank Urs Gasser for his advice and Peter Shane for comments on the first draft of this paper.

[1] *Viacom International, Inc. v. YouTube, Inc.*, No. 07 Civ. 2103. The case was later settled, which goes to show exactly how safe Google felt the safe harbor provisions to be. See Joe Silver. *Viacom and Google settle $1 billion YouTube lawsuit*. Available at http://arstechnica.com/tech-policy/2014/03/viacom-and-google-reach-settlement-in-long-running-youtube-lawsuit/ (last visited Aug 18 2014).

[2] Google Spain SL, Google Inc. v Agencia Española de Protección de Datos (AEPD), Mario Costeja González. Case C-131/12.

– be it social networks, video streaming websites, forums or peer-to-peer file-sharing networks. Rather, this article proposes a first approach on the subject. The driving purpose is to find a solution to the increasingly dire situation of online intermediaries – without which the internet as we know simply wouldn't exist.

II. Internet Users' Deep-Rooted Wish for Self-Governance

For many years the idea that behavior on the internet couldn't be regulated was very popular. It was a whole new world where the entities that exercised regulation either couldn't enter or did so only to remain at the same level as individual users. A court system, thousands of police officials, large armies, nuclear missiles, none of this mattered in the virtual world because it was inherently free and uncontrollable. Regulation by the "weary giants of flesh and steel" was not believed to be possible by internet users because governments have "no moral right to rule us nor [do they] possess any methods of enforcement we have true reason to fear."[3] That may very well have been true while the internet was still in its academic and hippie era. The 1990's, however, brought with them the phenomenon of the commodification of the internet.[4]

Notwithstanding the appropriateness of understanding the internet as a new and different place[5], the fact is that once it was noticed as a good forum for commercial activity private companies flocked to it. Their need for legal certainty and stability was a driving force in the alteration of technical standards that had earlier prevented the possibility of regulation. Changes effected in the net's architecture gradually enabled governments to exercise increased control to the point where the issue was no longer whether to regulate, but rather how to go about doing it. The fact that it constitutes a distinct place for human interaction doesn't automatically make it an isolated place: web

---

[3] John Perry Barlow, *A Declaration of the Independence of Cyberspace*. Available at: https://projects.eff.org/~barlow/Declaration-Final.html (last visited Apr 23 2012).

[4] This phenomenon has been described by many authors. See, for instance, Graham Murdock and Peter Golding's explanation: "Economicaly it involves moving the production and provision of communications and information services from the public sector to the market, both by transferring ownership of key facilities to private investors and by making success in the marketplace the major criterion for judging the performance of all communications and information organizations." Graham Murdock and Peter Golding, *Information poverty and political inequality*. *in*: The information society. v. III (Democracy, governance and regulation), 15 (Robin Mansell, org., 2009). This transition is achieved on the internet by prioritizing data flow based on merit attributed by market criteria: if streaming a movie makes more direct money than disseminating a post in a political blog, the latter is left with lower bandwidth. Howard Rheingold had predicted that once this transition is completed the internet will turn into a mass communication media not unlike cable television. "The great power of the idea of electronic democracy is that technical trends in communications technologies can help citizens break the monopoly on their attention that has been enjoyed by the powers behind the broadcast paradigm – the owners of television networks, newspaper syndicates, and publishing conglomerates.". Howard Rheingold, *The virtual community: homesteading on the electronic frontier* 308 (2000).

[5] "Cyberspace is a place. People live there. They experience all the sorts of things that they experience in real space, there. For some, they experience more. They experience this not as isolated individuals, playing some high tech computer game; they experience it in groups, in communities, among strangers, among people they come to know, and sometimes like." Lawrence Lessig, *The zones of cyberspace*, 48 Stan. L. Rev.1403, 1403 (1996). See also Colin Crawford, *Cyber*place*: defining a right to Internet access through public accommodation law*, 76 Temp. L. Rev. 225 (2003).Crawford

users are the same people who live within the borders of nation states and even those who don't access the internet are nonetheless affected by it. Total separation, although legally possible with the recognition of an independent cyberspace jurisdiction[6], is unpractical and unreal. The contention that it is impossible to track information flow online was perhaps partially true up until the mid-1990's. However, the use of Deep Packet Inspection[7] and other mechanisms has allowed internet service providers (ISPs) and governments to constantly and effectively control online communication. Another common argument was that it was impossible to identify the location of the people exchanging information on the internet. This difficulty was frequently posed in conjunction with that of the inconvenience of allowing one nation to enforce its laws upon citizens of other countries[8]. While geolocation software has all but solved the problem, the existence of conflicts involving the law of different countries was never something pioneered by the internet.[9]

Therefore, after a period of exhilarating freedom in an environment that was by its nature hostile to regulation, the internet was taken by commercial activity and had its technical rules changed just enough so as to adapt to the needs of private companies. For-profit websites covered the landscape and the cyberflâneur was gone.[10] Although there's a case to be made that such modifications to the internet architecture in order to solve the transborder law enforcement tribulations will mean a departure from the kind of

---

[6] "Many of the jurisdictional and substantive quandaries raised by bordercrossing electronic communications could be resolved by one simple principle: conceiving of Cyberspace as a distinct 'place' for purposes of legal analysis by recognizing a legally significant border between Cyberspace and the 'real world.'" David R. Johnson and David Post, *Law And Borders - The Rise of Law in Cyberspace*, 48 Stanford Law Review 1367, 1378 (1995). The authors do not, however, affirm that cyberspace and the physical world are perfectly separable.

[7] This technique allows an ISP to search into the data packets that carry information on the internet, thus searching, e.g., someone's email to check whether they have used a certain word. See Alex Wawro, *What Is Deep Packet Inspection?* Available at: http://www.pcworld.com/article/249137/what_is_deep_packet_inspection.html (last visited Apr 26 2012).

[8] Or even of one state being able to impose its laws on citizens of another state in a federalist national system. "The average user simply cannot afford the cost of defending multiple suits in multiple jurisdictions, or of complying with the regulatory requirements of every jurisdiction she might electronically touch. Thus, the need for dormant commerce nullification of state overreaching is greater on the Internet than any previous scenario." Dan L. Burk, *Federalism in Cyberspace*, 28 Conn. L. Rev., 1095, 1126 (1996).

[9] "They also are no more complex or challenging than similar issues presented by increasingly prevalent real-space events such as airplane crashes, mass torts, multistate insurance coverage, or multinational commercial transactions, all of which form the bread and butter of modern conflict of laws." Jack L. Goldsmith, *Against Cyberanarchy*, 65 University of Chicago Law Review 1199, 1234 (1998).

[10] "Something similar has happened to the Internet. Transcending its original playful identity, it's no longer a place for strolling — it's a place for getting things done. Hardly anyone "surfs" the Web anymore. The popularity of the "app paradigm," whereby dedicated mobile and tablet applications help us accomplish what we want without ever opening the browser or visiting the rest of the Internet, has made cyberflânerie less likely. That so much of today's online activity revolves around shopping — for virtual presents, for virtual pets, for virtual presents for virtual pets — hasn't helped either. Strolling through Groupon isn't as much fun as strolling through an arcade, online or off." Evgeny Morozov, *The Death of the Cyberflâneur*. Available at: http://www.nytimes.com/2012/02/05/opinion/sunday/the-death-of-the-cyberflaneur.html?pagewanted=all (last visited Apr 23 2012)

communication network whose potential was lauded as revolutionary[11], the fact remains that it's perfectly possible to regulate internet behavior and this has been done for years now.

A whole different issue is whether the internet should be regulated in the first place, especially by nation-states? Many of the current arguments for multistakeholderism in international internet governance[12] and self-regulation by the private sector are built on top of beliefs shared by many authors who in the late 1990's and early 2000's openly rejected government regulation of the net, even assuming that technically could be done. A common view was that state authority should be rejected as unnecessary: in a "cyberpopulist" model, "netizens" could themselves decide the rules that would govern them, adopting a direct democracy system. This idea has been dismissed by some as unrealistic and blind to the contribution of a representative legislating body that no society can do without,[13] but construed by others as a new justification for sovereignty: instead of a liberal state, power comes from the free choice of people to gather online in their self-governed communities.[14] A different proposed model was the recognition, by the nation state, of a new type of rulemaking process – one that isn't performed by government and is also (and perhaps because of that) internationally applicable. A *lex informatica* would be developed by repeated social practices online (customs) and by technical standards[15], accepting the decisive regulatory role played by choices on how the internet architecture is configured.[16] Unlike in the

[11] See Jonathan Zittrain, *Be Careful What You Ask For: Reconciling a Global Internet and Local Law*. Harvard Law School Harvard Law School Public Law Research Paper No. 60 (2003). Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=395300.

[12] Multistakeholderism is an approach to internet regulation that requires civil society to participate in decision-making along with governments and representatives of the private sector and academia. Among many proponents of such approach, see Milton Mueller et al., *The Internet and Global Governance: Principles and Norms for a New Regime*, 13 Global Governance 237, 250 (2007): "[M]ultistakeholder governance should be legitimized and maintained. This norm is a logical extension of principles relating to private networks and global scope. The Internet is in effect a global confederation of network operators and users and should not be regulated in a top-down manner via agreements among states alone." See also Wolfgang Kleinwächter, *Internet co-governance. Towards a multilayer multiplayer mechanism of consultation, coordination and cooperation (M3C3)*, in: The information society. v. III (Democracy, governance and regulation), 384 (Robin Mansell, org., 2009).

[13] "First, cyberpopulists overestimate the extent to which the plebiscite, whether territorial or virtual, can truly reflect the voice of the people. Second, they ignore significant democracy-enhancing benefits of representative government." Neil Weinstock Netanel, *Cyberspace Self-Governance: A Skeptical View from Liberal Democratic Theory*, 88 California Law Review 395, 417 (2000).

[14] David Post credits the choice of self-government and free association online with the possibility of acknowledging sovereignty to internet users. This would be an alternative to the liberal state theory of sovereignty, where the agents of power and decision-making capacity are netizens themselves. David G. Post, *„The Unsettled Paradox": The Internet, The State, and the consent of the Governed*, 5 Ind. J. Global Legal Stud. 512, 535-539 and 542 (1998).

[15] "The source of default rules for a legal regime is typically the state. The political-governance process ordinarily establishes the substantive law of the land. For Lex Informatica, however, the primary source of default rule-making is the technology developer and the social process by which customary uses evolve." Joel R. Reidenberg, *Lex Informatica: The Formulation of Information Policy Rules Through Technology*, 76 Texas Law Review 553, 571 (1998).

[16] This is the landmark contribution of Lawrence Lessig to the field of internet regulation. The way the code is written creates a constraint on action online just as much as law does on action offline. The key difference, however, is that regulation by code is by its nature ex ante, whereas law is ex post facto – the former prevents an individual from even doing something in the first place; the latter punishes certain

cyberpopulist model, *lex informatica* would be enforced by government, such that the latter would merely loose its rulemaking prerogative[17], and even then only on what concerns human action online.[18] Even those who accepted enforcement of traditional legal norms, especially regarding commerce on websites, argued for concessions. A company couldn't be considered to be offering its products or services to everyone in the whole world. As adjudication of online conflicts slowly developed, it seemed reasonable to recognize that companies often targeted a specific audience despite the fact that their website was viewable to anyone.[19]

It's very important to notice that both models fundamentally evoke self-government, just like advocates of the impossibility of regulating internet did. John Perry Barlow's 1996 Declaration of Independence of Cyberspace symbolized a view that was more about autonomy of internet users to establish their own rules then about the technical impossibility of state regulation of the internet. But in order for this governance model to even have a shot at succeeding, a delicate balance must be struck between the people's freedom to leave a community whenever they so desire and, on the other hand, a reason for them to stay that is strong enough to maintain some stability in the composition of the community over time.[20]

The point is that there has been great force, for many years, in the idea that internet users deserve a higher level of autonomy to make and indeed enforce their own rules regarding online conduct. This arguably derives from a notion that states are not well suited to make regulatory decisions concerning the internet because the traditional state decision-making mechanisms and actors completely fail to grasp the reality of the

---

behavior after it has been engaged on. Lawrence Lessig, *Code: And Other Laws of Cyberspace*, Version 2.0, 7 (2006). An important fact that should not be overlooked is that law constrains human conduct directly, ex post facto, and indirectly, by influencing the code or architecture of the internet itself. See Lawrence Lessig, *The New Chicago School*, 27 The Journal of Legal Studies 661, 666 (1998).

[17] That is because "[*l*]*ex Informatica* has three sets of characteristics that are particularly valuable for establishing information policy and rule-making in an Information Society. First, technological rules do not rely on national borders. Second, *Lex Informatica* allows easy customization of rules with a variety of technical mechanisms. Finally, technological rules may also benefit from built-in self-enforcement and compliance-monitoring capabilities." Reidenberg, *supra* note 13, at 577.

[18] A much less romantic view is that this is none other than a free market mechanism for regulation of conduct, such that it is "essential to permit the participants in this evolving world to make their own decisions. That means three things: make rules clear; create property rights where now there are none; and facilitate de formation of bargaining institutions. Then let the world of cyberspace evolve as it will, and enjoy the benefits." Frank H. Easterbrook, *Cyberspace and the law of the horse*, 1996 U. Chi. Legal F. 207, 216 (1996). The problem is, of course, that creating property rights invites more, rather than less, state intromission as it is government that protects individual property through private law rules of contract and civil liability.

[19] "Courts have almost universally required some additional proof of either traditional commercial contacts or intentional direction of the activity toward the forum – a form of purposeful availment. Because some courts have allowed plaintiffs to conduct "jurisdictional discovery" and have also occasionally found the web site's records of forum visitors to be relevant, it would seem prudent for states seeking to enforce their laws against outlaw websites to seek discovery of the web server logs in order to attempt to make a sufficient record as to the number of forum contacts." Terrence Berg, *www.wildwest.gov: The impact of the Internet on state power to enforce the law*, 2000 BYU L. Rev. 1305, 1338 (2000).

[20] "[W]hen individuals have a substantial stake in a particular virtual community, exit is not a tenable option to protect them against majority oppression. But when individuals lack that investment, the result is a flame-ridden cacophony rather than a cohesive community capable of government by the "bottom-up" generation of social norms". Netanel, *supra* note 11, at 432.

internet. As a result, internet users are often eager to take regulation into their own hands. They feel empowered, in control, and most importantly, legitimated to create and apply rules and principles on behavior. This is different from social norms, which are created by a practice repeated over a long time, engendering a social custom. Some of the rules internet users obey in their communities are of that kind, but others are expressly, voluntarily created and codified, much like legal norms.[21] It's obvious that these two types of self-imposed rules have an intrinsic relationship such as that of law and morals[22] and therefore an attempt at a clear split would be both unwise and difficult. My point is merely that while internet users often recognize the force of both, the codified, written rules that they spontaneously create undeniably demonstrate an assertion of self-governance prerogative. Under the appropriate circumstances, this enthusiasm could be harnessed by a private company managing an online community.

III. The Problems Faced by Online Intermediaries

Commercial web pages and online applications currently thrive whenever they can establish and sell themselves as a platform. Very few internet startups incorporate to their business plan the autonomous production of content. What they expect is to create an environment where social interaction based on the contributions of users themselves would boost the popularity of their platform[23]. The community sentiment is stimulated not only to motivate users to create content, but to suggest the impression of a shared commons, where users feel that they are voluntarily collaborating for a mutual purpose and that each of them has a stake in the continuation of the platform[24]. Autonomy and self-governance are a decisive part of this sentiment.

This focus by internet companies to play the role of an intermediary instead of the content producer has raised, along with the activity of internet service providers, the hotly debated question of online intermediary liability. This is arguably one of the key issues in the developing field of cyberlaw in several countries for at least three reasons.

First, it involves the assertion of legal responsibility that will trigger compensation and thus large sums of money are awarded according to how the questions are answered. Second, it deals with issues relating to the technical design of platforms such as social networks, search engines, web 2.0 applications, p2p software and any website that invites user-generated content. Third, and perhaps most importantly, it serves as a backdrop against which the protection of fundamental rights like freedom of expression, privacy, property and freedom to conduct a business are balanced.

Online intermediaries all have to face a dire and pressing matter: *will* they filter and censor content created by their users / customers? *Could* they engage in such filtering? *Should* they? Will they be liable when users in their platform violate the privacy or property of third parties? In the mid-90's this was noticed as an entirely new and

[21] I'm adopting the distinction between customary norms and legal (positive) norms made by Hans Kelsen, *Pure Theory of Law* (2nd ed., 1978).

[22] As described by H. L. A. Hart, *Positivism and the Separation of Law and Morals*, 71 Harvard Law Review (1958).

[23] See Jeff Jarvis, *What Would Google Do?* (2009).

[24] Companies like Zipcar currently tap into this inherent selflessness of humans as a way to strengthen their business. See Yochai Benkler, *The Penguin and the Leviathan: How Cooperation Triumphs over Self-Interest* (2011).

incredibly relevant issue because never had companies relied so heavily and successfully on the contribution of customers for the operation of their business, while at the same time foregoing the exercise of an editorial function whereby the managers go through all of the content produced or shared by users. They profited from the input of customers, but they weren't exercising review. This was a defining moment for online crowdsourcing and had the United States (the country where the absolute majority of such innovative companies are settled and where most of the users originate) opted for attributing liability to the intermediaries this industry as we know it today arguably wouldn't exist[25]. But the solution found was to give immunity to intermediaries when users exchange data that infringes copyright[26] or constitutes lewd speech[27]. In theory, this would have meant that internet companies were saved from a big headache and could further conduct their business unhampered by fear of liability. But the actual picture is much different.

Firstly, the adoption of a safe harbor for intermediaries in the United States wasn't followed by the same choice in all other countries. Fortunately, the European Union's e-commerce directive, enacted in 2000, mandated member states to ensure that intermediaries would not be held liable[28], similarly to what had been done by American

---

[25] This is why the Digital Millennium Copyright Act of 1998 has been hailed as the law that saved the internet. David Kravets, *10 Years Later, Misunderstood DMCA is the Law That Saved the Web.* Available at: http://www.wired.com/threatlevel/2008/10/ten-years-later/ (last visited Apr 26 2012).

[26] Section 512 of the Digital Millennium Copyright Act of 1998 reads: "(a) TRANSITORY DIGITAL NETWORK COMMUNICATIONS- A service provider shall not be liable for monetary relief, or, except as provided in subsection (j), for injunctive or other equitable relief, for infringement of copyright by reason of the provider's transmitting, routing, or providing connections for, material through a system or network controlled or operated by or for the service provider, or by reason of the intermediate and transient storage of that material in the course of such transmitting, routing, or providing connections, if (1) the transmission of the material was initiated by or at the direction of a person other than the service provider; (2) the transmission, routing, provision of connections, or storage is carried out through an automatic technical process without selection of the material by the service provider; (3) the service provider does not select the recipients of the material except as an automatic response to the request of another person; (4) no copy of the material made by the service provider in the course of such intermediate or transient storage is maintained on the system or network in a manner ordinarily accessible to anyone other than anticipated recipients, and no such copy is maintained on the system or network in a manner ordinarily accessible to such anticipated recipients for a longer period than is reasonably necessary for the transmission, routing, or provision of connections; and (5) the material is transmitted through the system or network without modification of its content."

[27] Section 230 of the Communications Decency Act reads: "No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider."

[28] 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. Article 15 (1): "Member States shall not impose a general obligation on providers, when providing the services covered by Articles 12, 13 and 14, to monitor the information which they transmit or store, nor a general obligation actively to seek facts or circumstances indicating illegal activity." The Directive left open the possibility that legislation would provide injunctive relief for copyright holders against intermediaries in order to cease infringement, but not to obtain compensation. Four years later, the Directive on intellectual property rights required that such and injunction be made available for judicial authorities in member states. Directive 2004/48/EC of The European Parliament and of The Council of 29 April 2004 on the enforcement of intellectual property rights", Article 9 (Provisional and precautionary measures) (1) Member States shall ensure that the judicial authorities may, at the request of the applicant: (a) (…) an interlocutory injunction may also be issued, under the same conditions, against an intermediary whose

legislation. This has proved to be a not-so-safe harbor for companies in Europe. In 2010, Google executives themselves were criminally convicted in Italy of privacy invasion due to a video that was posted of a boy with autism being beaten by other boys[29]. This shows that even if legislation grants intermediaries immunity for copyright violations, there are still other illegal uses of user-generated platforms that might warrant liability according to the national legal system where a global company such as Google or Facebook is operating.

The copyright industry has been the greatest champion of intermediary liability. The Belgian Society of Authors, Composers and Publishers (SABAM) has twice tried and twice failed, within a short interval, to obtain a ruling by the European Court of Justice that would impose on internet intermediaries the obligation to monitor information flow between users. The decision issued on November, 2011 denied that ISPs could be legally forced to monitor copyright infringement by their customers[30]. The one issued on February, 2012 confirmed its predecessor, now exempting online social network operators from a duty to filter content in order to block copyright infringing material[31]. In both cases the reasoning of the Court was that imposing an absolute blanket-censorship obligation on ISPs and social networks was a disproportional balancing of the rights to receive and impart information, to privacy, and to conduct a business activity, on one hand; and to (intellectual) property on the other. SABAM's success in taking these cases all the way up to the ECJ 12 years after the safe harbor rule was enshrined in the e-commerce Directive illustrates the constant liability threat under which platform providers find themselves in Europe. Furthermore, it shows that even if the law has established the absence of liability, intermediaries have a perpetual disbursement of resources in order to pay for litigation costs.

The ECJ's "right to be forgotten" ruling in May 2014 drives the point home. Privacy protection was understood to trump safe harbor or at least call for a different, lest protective interpretation of it. Google was ordered to filter search results regardless of specific court orders. That creates precisely the level of risk and uncertainty[32] for the intermediary that the safe harbor rule was enacted to prevent. Even legislators in Europe, who have been quarreling with American tech giants in the past few years over tax evasion allegations[33], came out against the ruling stating it is "unworkable"[34]. That is

---

services are being used by a third party to infringe an intellectual property right; injunctions against intermediaries whose services are used by a third party to infringe a copyright or a related right are covered by Directive 2001/29/EC."

[29] The video had been posted to the Google Video website in 2006. See Stephen Shankland, *Execs convicted in Google Video case in Italy*. Available at: http://news.cnet.com/8301-30685_3-20000092-264.html?tag=newsEditorsPicksArea.0 (last visited Apr 24 2012).

[30] Scarlet Extended SA v Société belge des auteurs, compositeurs et éditeurs SCRL (SABAM). Case C-70/10.

[31] Société belge des auteurs, compositeurs et éditeurs SCRL (SABAM) v Netlog NV. Case C-360/10. SABAM's strategy was to interpret the IP Directive of 2004's guarantee of injunction against intermediaries to cease infringement as a right to force them to implement and maintain, at their own expense, a permanent filtering system.

[32] One need only take a quick look at paragraph 99 of the ruling to grasp how subjective supposed standard is and how problematic it will be for any intermediary – even with the help of legal counsel – to make filtering decisions based on it. C-131/12, supra note 2.

[33] See Andrew Frye. *Renzi Pressed to Put Google, Facebook Taxes on EU Agenda*. Available at http://www.bloomberg.com/news/2014-07-01/renzi-pressed-to-put-google-facebook-taxes-on-eu-

because "[t]he requests received in June alone mean that Google's staff have to review over a quarter of a million URLs to see whether the information appears to be "inadequate, irrelevant or no longer relevant, or excessive in relation to the purposes of the processing" carried out by them." [35] The future for intermediaries in Europe is indeed grim.

The prospect in other countries is shaky. In 2014 Brazil enacted its *Marco Civil da Internet*, a groundbreaking landmark internet regulation statute[36] that contains several provisions regarding intermediaries. Instead of notice-and-takedown, the system adopted was court-order-and-takedown. While this is good news to intermediaries, copyright violations and child pornography accusations were left out of this strong safe harbor and tend to be solved by Brazilian courts with notice-and-takedown or something even worst. And the Judiciary's track record is certainly a bad omen.

Provisions of the Consumer Protection Code on strict liability of service providers who engage in risky activity have been often interpreted as requiring liability of social networks for defamation engaged in by its users. In 2010 the district attorney's office of the state of Rondônia filed a civil suit against Google due to the existence of several communities created within the company's Orkut social network where teenagers were the target of libelous gossip. An injunction was asked and granted, whereby in addition to delete all illegal content and identify the creators of such communities, Google was ordered to prevent the creation of similar communities. The companied denied complying with this last command and an appeal reached the Brazilian Superior Federal Court of Justice. In deciding whether or not Google could be legally mandated to exercise prior restraint on content created by users in its social network, the Justice referred that the company's argument of technical impossibility of prior censorship in an online crowd-sourced platform was without merit, given that the Chinese government executed online censorship all the time[37]. The Court reversed this precedent in the following year when it

---

agenda.html (last visited Aug 18 2014); Frances Robinson. *France Pushes EU to Regulate U.S. Internet Companies*. Available at http://online.wsj.com/news/articles/SB10001424127887324492604579085222987377040 (last visited Aug 18 2014). See also House of Commons. Committee of Public Accounts. *Tax Avoidance–Google*. Ninth Report of Session 2013–14. Available at http://www.publications.parliament.uk/pa/cm201314/cmselect/cmpubacc/112/112.pdf (last visited Aug 18 2014), which reports that in order "[t]o avoid UK corporation tax, Google relies on the deeply unconvincing argument that its sales to UK clients take place in Ireland, despite clear evidence that the vast majority of sales activity takes place in the UK."

[34] Alex Hern. Lords describe Right to be Forgotten as 'unworkable, unreasonable, and wrong'. Available at http://www.theguardian.com/technology/2014/jul/30/lords-right-to-be-forgotten-ruling-unworkable (last visited Aug 18 2014). Ong decisions based on it. C-131/12, supra note 2.

[35] European Union Committee. Second Report. *EU Data Protection law: a 'right to be forgotten'?* Available at http://www.publications.parliament.uk/pa/ld201415/ldselect/ldeucom/40/4002.htm (last visited Aug 18 2014), at paragraph 33.

[36] Glyn Moody. *Brazil's 'Marco Civil' Internet Civil Rights Law Finally Passes, With Key Protections Largely Intact*. Available at https://www.techdirt.com/articles/20140326/09012226690/brazils-marco-civil-internet-civil-rights-law-finally-passes-with-key-protections-largely-intact.shtml (last visited Aug 18 2014).

[37] RECURSO ESPECIAL Nº 1.117.633 - RO (2009/0026654-2). "O provedor de serviços responsável pela manutenção do orkut já se utiliza da fiscalização de conteúdo em outros países, como é o caso da China, não sendo possível vislumbrar, de início, em que a situação ora analisada difere da que vem sendo empregada naquele país." ("The service provider responsible for maintaining Orkut already employs content monitoring in other countries, as is the case of China, which is why it is not possible to see,

decided another appeal on the merits of a similar case[38]. The Supreme Constitutional Court has picked up a similar case for judgment in the upcoming months and could go one way or the other. In theory it isn't even bound by the *Marco Civil* choice for court-order-and-takedown because the Justices could rule that the Constitution requires more effective protection for defamation victims.

The second reason why companies can't completely ignore the content of exchanges in their platforms is that the safe harbor rule has caveats and the result of judicial interpretation over the last ten years in the United States hasn't been entirely favorable to intermediaries. Companies have to find a sweet spot between managing their online platform to achieve their business goals and avoiding a level of intervention on the activity of users that would characterize editorial action and thus trigger liability. This has been the case for p2p software, where since Napster the developers need to centralize control of the file exchange process enough to be able to coordinate it and keep things running smoothly, but at the same time have to maintain the level o centralized coordination below a certain threshold over which indirect liability ensues[39].

Two cases that reached federal appeals courts show that the distinction between a liable intermediary and one that is in safe harbor is workable but by no means a clear-cut rule. This invites case-by-case interpretation and therefore keeps the possibility of finding the intermediary liable always present. In Fair Housing Council of San Fernando Valley v. Roommates.com, LLC, 521 F.3d 1157 (9th Cir. 2008), the manager of a website that served as a platform where anyone could find other people to share an apartment with was deemed liable for discrimination under the Fair Housing Act. Individuals who had to participate were forced to fill out a profile which asked for information on gender, sexual orientation and number of children, among other personal information. The website's search engine featured filtering options that employed these criteria. In order to determine whether Roommates.com was an interactive computer service (immune) or an information content provider (liable), the Court asserted whether the platform manager

---

initially, how the situation under analysis here is different from the one used in that country.") (author's translation). On this ruling, as well as on the protection of free speech online in Brazil on a more general matter, see Sam Bayard, *Brazil Fines Google Over Dirty Jokes on Orkut; Brazilian Lawyers Weigh In*. Available at: http://www.citmedialaw.org/blog/2010/brazil-fines-google-over-dirty-jokes-orkut-brazilian-lawyers-weigh (last visited Apr 24 2012).

[38] RECURSO ESPECIAL Nº 1.193.764 - SP (2010/0084512-0). Justice Nancy Andrighi concluded that "não se pode considerar de risco a atividade desenvolvida pelos provedores de conteúdo, tampouco se pode ter por defeituosa a ausência de fiscalização prévia das informações inseridas por terceiros no site, inexistindo justificativa para a sua responsabilização objetiva pela veiculação de mensagens de teor ofensivo." ("the activity developed by content providers cannot be considered as a risky one; the absence of prior restraint on information inserted in the website by third parties also cannot be seen as a defect in the service, therefore remaining without justification a strict liability of such providers for messages of offensive content made available on the website") (author's translation). The opinion explicitly finds support on the DMCA safe harbor provision as well as on the European e-commerce Directive. Doctrine in Brazil is generally more receptive to a safe harbor system than to strict liability. See Marcel Leonardi, *Responsabilidade Civil dos Provedores de Serviços de Internet* (2005) and Bruno Miragem, *Responsabilidade por danos na sociedade de informação e proteção do consumidor: desafios atuais da regulação jurídica da internet*, 70 Revista de Direito do Consumidor 41 (2009).

[39] See Tim Wu, *When code isn't law*, 89 Va. L. Rev. 679, 724 (2003): Wu points to the sense of community that exists between users of p2p systems as one of the reasons for the success of such platforms.

acted as a content co-developer and whether it had induced infringement[40]. Both questions were answered in the affirmative for Roommates.com, which was therefore found liable. The result was different in Chicago Lawyers' Committee For Civil Rights Under Law, Inc. v. Craigslist, Inc., 519 F.3d 666 (7th Cir., 2008). Challenged under the same Fair Housing Act accusation, that it was liable for discriminatory housing ads posted by its users, Craigslist was granted immunity because the Court felt it didn't in any way induce users to post such ads – it had no mandatory boxes that a user had to fill-in in order to use the platform[41]. But Judge Easterbrook explicitly denied that safe harbor could work as a rule that would give clear safety to intermediaries if they chose not to worry about the content or messages exchanged by their users. In relying on a Supreme Court precedent, he stated: "[t]o appreciate the limited role of § 230(c)(1), remember that "information content providers" may be liable for contributory infringement if their system is designed to help people steal music or other material in copyright. See Metro-Goldwyn-Mayer Studios Inc. v. Grokster, Ltd., 545 U.S. 913, 125 S. Ct. 2764, 162 L. Ed. 2d 781 (2005); In re Aimster Copyright Litigation, 334 F.3d 643 (7th Cir. 2003). Grokster is incompatible with treating § 230(c)(1) as a grant of comprehensive immunity from civil liability for content provided by a third party."[42]

Grokster wasn't about a social network or another type of website, rather it concerned a p2p software. This shows that intermediary liability is an overarching issue encompassing any platform operator that employs the internet to interconnect individuals and let them exchange information of any kind – pictures, status updates, comments, music files etc. Metro-Goldwyn-Mayer Studios Inc. v. Grokster, Ltd., 545 U.S. 913 (2005) marks a strong shift away from anything resembling a safe harbor for platform providers. In Grokster the Court went beyond the standard it had set in Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417 (1984). While in Sony the existence of non-infringing uses of a technology ensured that the developer would not be held liable[43], in Grokster the Court ventured into the intentions of the platform developer. Regardless

---

[40] "CDA does not grant immunity for inducing third parties to express illegal preferences. Roommate's own acts--posting the questionnaire and requiring answers to it--are entirely its doing and thus section 230 of the CDA does not apply to them. Roommate is entitled to no immunity." Fair Housing Council of San Fernando Valley v. Roommates.com, LLC, 521 F.3d at 1165. Judge Kozinski's opinion also affirmed "that reading the exception for co-developers as applying only to content that originates entirely with the website--as the dissent would seem to suggest--ignores the words "development . . . in part" in the statutory passage "creation or development in whole or in part." 47 U.S.C. § 230(f)(3) (emphasis added). We believe that both the immunity for passive conduits and the exception for co-developers must be given their proper scope and, to that end, we interpret the term "development" as referring not merely to augmenting the content generally, but to materially contributing to its alleged unlawfulness. In other words, a website helps to develop unlawful content, and thus falls within the exception to section 230, if it contributes materially to the alleged illegality of the conduct." Id., at 1167-1168.

[41] Craigslist was understood as a common carrier in this sense: "Online services are in some respects like the classified pages of newspapers, but in others they operate like common carriers such as telephone services, which are unaffected by § 3604(c) because they neither make nor publish any discriminatory advertisement, text message, or conversation that may pass over their networks." Chicago Lawyers' Committee For Civil Rights Under Law, Inc. v. Craigslist, Inc., 519 F.3d at 668.

[42] Id., at 670.

[43] "the sale of copying equipment, like the sale of other articles of commerce, does not constitute contributory infringement if the product is widely used for legitimate, unobjectionable purposes. Indeed, it need merely be capable of substantial noninfringing uses." Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. at 442.

of the possibility of using the platform for legal purposes, the manager could be held liable if it had induced infringement[44]. Of course Justice Souter in Grokster did all he could to make it seem as though the Sony rule wasn't being abandoned, but the fact is the rule changed[45], making intermediary liability more uncertain than it was before the ruling[46]. This was only the culmination of an ongoing process in lower courts, where the changes being made to the law were increasing the uncertainty for platform developers[47].

Safe harbor for online intermediaries has thus been turned into an indirect liability standard, one which inevitably curtails legitimate use that individuals may make of a legitimate online platform[48]. Furthermore, advocates of the fight against online child pornography call for a revision of the safe harbor provision in the CDA[49]. The only

---

[44] "where evidence goes beyond a product's characteristics or the knowledge that it may be put to infringing uses, and shows statements or actions directed to promoting infringement, Sony's staple-article rule will not preclude liability." Metro-Goldwyn-Mayer Studios Inc. v. Grokster, Ltd., 545 U.S. at 935.

[45] "The Court created a new type of contributory copyright infringement—while apparently denying it was doing so." James Boyle, *The Public Domain. Enclosing The Commons of the Mind* 77 (2008).

[46] "[T]here is no such thing as a bright-line rule for technologists to make reliable ex ante determinations as to what it means to be too close to the line of secondary copyright liability in the Post-Grokster World." Urs Gasser and John G. Palfrey, Jr., *Catch-As-Catch-Can: A Case Note on Grokster* Research Publication No. 2005- October 2005, at 14. This uncertainty is more than enough to hamper online platform providers: "A decision does not need to make an activity illegal in order to impede it. It only needs to make it uncertain." Boyle, supra note --, at 79.

[47] Jonathan Zittrain points out that in In re Aimster Copyright Litig., 334 F.3d 643 (7th Cir. 2003) "[t]he Seventh Circuit's test put all authors of generative technologies at risk of finding themselves on the wrong side of a court's cost/benefit balancing. Indeed, they were asked to actively anticipate misuses of their products and to code to avoid them. Such gatekeeping is nice when it works, but it imposes extraordinary costs not readily captured by a single cost/benefit test in a given instance." Jonathan Zittrain, *A History of Online Gatekeeping*, 19 Harv. J.L. & Tech. 253, 285 (2006).

[48] "Indirect liability has a significant drawback, however, in that legal liability — even if carefully tailored — inevitably interferes with the legitimate use of implicated tools, services, and venues. (…)This concern is particularly pronounced for new technologies, where the implications of copyright liability are often difficult to predict." Douglas Lichtman and William Landes, *Indirect Liability For Copyright Infringement: An Economic Perspective*, 16 Harv. J.L. & Tech. 395, 409 (2003). That is because in countries like the United States, the main driving force of intermediary liability online is copyright protection, something which is inherently in tension with the protection of freedom of expression, since "[r]ecognizing property rights in information consists in preventing some people from using or communicating information under certain circumstances. To this extent, all property rights in information conflict with the "make no law" injunction of the First Amendment." Yochai Benkler, *Free as The Air to Common Use: First Amendment Constraints on Enclosure of The Public Domain*, 74 N.Y.U. L. Rev. 354, 393 (1998).

[49] The shaky indirect liability standard would then be replaced by a free-for-all negligence standard: "The young person (or his parents, more likely, I suppose) seeks to bring suit against the service provider involved. In my view, the service provider should not have special protection from such a tort claim. Such a claim should be decided on the merits. Was the service provider negligent or not? I don't think that the fact that the service provider is offering an Internet-based service, rather than a physically based service, should result in a shield to liability." John Palfrey Jr. *in* Adam Thierer, *Dialogue: the future of online obscenity and social networks*. Available at: http://arstechnica.com/tech-policy/news/2009/03/a-friendly-exchange-about-the-future-of-online-liability.ars (last visited Apr 24 2012). Intermediary liability for child pornography involves a balancing of free speech with the need to protect vulnerable internet users – children – "who do not have the same skills as adults to make a broad range of quality judgments that accompany these informational processes – limitations that are due to their respective stage of development and their limited set of life experience based on which content can be evaluated." Urs Gasser et al., *Response to FCC Notice of Inquiry 09-94. Empowering Parents and Protecting Children in an Evolving Media Landscape*. Available at: http://ssrn.com/abstract=1559208 (last visited Apr 24 2012), at 3.

"safe" thing about all of this is that it's safe to say intermediaries cannot easily forego some kind of management of the content exchanged in the platform that they operate.

This carries its own set of problems, of course. First of all, engaging in filtering has a cost. It is commonly alluded to the impracticability of exercising human oversight over the activity in online platforms. The company-staff-to-user-base ratio makes individual examination of each exchange impossible. Automated filtering is by far not unfeasible, and filters like the ones used against spam by Gmail and against copyright infringement on Youtube employ sophisticated algorithms capable of lowering the so-called false negatives and false positives. They still have a cost, however, and even 5% of false positives represent a significant social cost when freedom of expression is involved. Second, if companies take it upon themselves to exercise the filtering that would keep them free of liability, there will be a natural tendency to filter more, not less[50]. Liability poses a big financial threat, one that isn't always offset by the dissatisfaction of a couple of users who had their posting, comment or video deleted. Third, the very possibility of user insurrection against filtering executed by the platform manager works as a force opposing that of risk-aversive over-censorship. Indeed, the intermediary finds itself between a rock and a hard place: if it filters content, consumers potentially react badly, organize and protest[51]; if it doesn't filter, it highly increases the chances of being held liable – and sometimes incurring in millionaire penalties. Litigation costs also need to be added to that bill. By 2010, four years before the settlement, Google had reportedly already spent U$ 100 million in legal fees with Viacom v. Youtube[52]. Actively filtering content in social networks creates a very bad image these days, even if the company explains that it does so in order to comply with government regulation[53].

IV. The Alternative – Harnessing User Self-Governance for Platform Management

[50] This predisposition for censorship amounts to a serious chilling effect on free speech. This problem has been described in detail by Wendy Seltzer, *Free Speech Unmoored in Copyright's Safe Harbor: Chilling Effects of The DMCA on The First Amendment*, 24 Harv. J.L. & Tech. 171 (2010). The author diagnosed the problem in a system of strong safe harbor, but warns of an even worst scenario under the relative intermediary liability standard: "Moreover, the chilling effect analysis indicates that over-deterrence is a problem deeper than the DMCA notice-and-takedown regime; it is a problem endemic to copyright law and its secondary liabilities. As copyright expands in scope, time, and breadth, its erroneous application and the chill of secondary liability assume greater significance." Id., at 227.

[51] Yochai Benkler explains how these seemingly decentralized, bottom-up user campaigns in a networked-society are effective at bending the will of powerful actors, including large private companies. One of the examples reported by Benkler is that of how users successfully forced Sinclair Broadcasting not to air a controversial TV ad during the 2004 presidential elections in the United States. See *The Wealth of Networks: How Social Production Transforms Markets and Freedom* 220 (2006).

[52] Liz Miller. *Google's Viacom Suit Legal Fees: $100 Million*. Available at http://gigaom.com/2010/07/15/googles-viacom-suit-legal-fees-100-million/ (last visited Aug 18 2014). While this only hurts a company like Google, it does much worse to startups. According to the story, litigation costs were largely responsible for Veoh's bankruptcy.

[53] A good example of this is the surge of criticism that followed Twitter's announcement that it would start to suppress tweets that governments in countries like Syria or Iran asked to be removed. See Somini Sengupta, *Censoring of Tweets Sets Off #Outrage*. Available at: http://www.nytimes.com/2012/01/28/technology/when-twitter-blocks-tweets-its-outrage.html?pagewanted=all (last visited Apr 25 2012). To its credit, the company has taken an unusual path and decided to maintain the transparency of the tweet removals, by acknowledging them explicitly in each case.

The ideal of self-governance by internet users was very poignant in the late 1990's and early 2000's. However momentum this push has lost due to the realization that countries can and do regulate the internet, it has not been fully abandoned. Users of many online platforms are willing to assert some level of self-governance prerogative whereby they perform norm-enforcing roles. This rises as an alternative for intermediaries: when giving up filtering for illegal content is risky and taking up filtering has financial and political costs, outsourcing this task to users themselves could potentially solve many of the platform manager's problems.

People sharing an environment over a certain time tend to cultivate a bond with the platform itself but also with the individuals that co-exist with them. This is the case in social networks like Facebook, user-moderated news websites like Slashdot, user-generated content websites like 9GAG or Wikipedia and virtual worlds like World of Warcraft[54]. A notion of community develops which evokes that independence and group self-determination feeling that has existed on the internet since the very beginning. It seems that the sense of communion is proportional to the level of detachment from the real world that the environment produces. In massively multiplayer online games this reaches perhaps the strongest stance [55]. In most other platforms maintained by intermediaries, however, user zeal for the common space and resources is pervasive and persistent. There is a big difference in the perception of users between the filtering enforced by the intermediary and that which is carried out by users themselves. The former is a bottom-down imposition of values; the latter is a bottom-up exercise of self-regulation and independent authority. While it is true that this authority derives from the desire of the intermediary to maintain a platform that is compliant with the law, and therefore not all values are necessarily shared by the company and its customers, to the extent that some are and users – not the company – take the leading role in putting them into practice, there are elements of self-governance to be found in this context. This greatly reduces the rejection and dissatisfaction by users with the filtering that is performed.

The issue is whether users could successfully entertain a task of governance. For the purposes of the analysis here, filtering for illegal content such as copyright infringing goods or child pornography will be the governance aspect considered. Censorship of this fashion would require, to some extent, shunning the users who engage in such

---

[54] On the potential of community filtering, Yochai Benkler states that "[c]onsistent with what we have been seeing in more structured peer-production projects like Wikipedia, Slashdot, or free software, communities of interest use clustering and mutual pointing to peer produce the basic filtering mechanism necessary for the public sphere to be effective and avoid being drowned in the din of the crowd." Benkler, *supra* note 44, at 258.

[55] The bonding and community-forming goals can be noticed in all kinds of online games, not only those such as Second Life: "As these examples indicate, each virtual world is different, making categorical statements about virtual worlds suspect. Still, the lines drawn between worlds might not be as bright as they seem at first. For instance, while *The Sims Online* does not involve gaining power and wealth through leveling, prestige and affluence are motivating forces for many participants. And while leveling worlds such as *Ultima Online* often force players to engage in repetitive killing exercises, what makes this bearable seems to be the social bonds formed among players, who may find more fulfillment in being virtual seamstresses, alchemists, and blacksmiths." F. Gregory Lastowka and Dan Hunter, *Virtual Worlds: A Primer*, *in* The State of Play. Law, Games, and Virtual Worlds 24 (Jack Balkin and Beth Noveck eds., 2006).

wrongdoings. This is clearly about the decentralized, commons-based production of an information service: filtering content. This model of production depends on modularity, granularity and heterogeneity[56]. The task of filtering can only be undertaken by users in a decentralized approach if the overall work can be broken up into pieces; if these pieces or isolated parts of the job are small; and if they are of different sizes and levels of complexity. It seems mechanisms such as red-flagging, which today are familiar to users of many social networks[57], go a long way in providing for modularity and granularity. Heterogeneity seems to characterize the task as well: while certain content is more obviously infringing than others, there are also those grey-area instances. There is the sale of copyrighted music and then there's remixing under fair use; there are pictures of naked children and then there are artistic paintings which include nude children among other elements.

User filtering is a mechanism of gatekeeping because it concerns the control of information flow[58]. At the same time, this is a peculiar kind of gatekeeping because it involves the traditional gated becoming the gatekeepers[59], the decision-makers on the issue of whether or not certain content passes scrutiny and can be shared in a community. This is decentralized gatekeeping whereby the users of an online platform purport to collectively fulfill a goal related to the control of information flow[60]. In order for user filtering to work, coordination isn't as essential as in other collective endeavors like the production of encyclopedia articles in Wikipedia[61]. It is nonetheless crucial that the users exchange their views or produce standards and general guidelines for the filtering, lest the whole process collapses with excessive or insufficient censorship. This doesn't mean that without unanimity on a general set of rules the whole enterprise is doomed to fail. Rough consensus can play a part in online communities[62], but in reality the existence of some

---

[56] Yochai Benkler, *Coase's Penguin, or, Linux and The Nature of the Firm*, 112 Yale L.J. 369, 435-436 (2002).

[57] In some websites this is even required of users. In Craigslist, for example, when people enter any of the subsections of the personal ads portion of the listings, they are prompted to agree to certain conditions in order to continue. One of them reads: "I agree to flag as "prohibited" anything illegal or in violation of the craigslist terms of use." Available at: http://boston.craigslist.org/cgi-bin/personals.cgi?category=stp (last visited Apr 26 2012).

[58] Gatekeeping can be defined "as the process of controlling information as it moves through a gate. Activities include, among others, selection, addition, withholding, display, channeling, shaping, manipulation, repetition, timing, localization, integration, disregard, and deletion of information." Karine Barzilai-Nahon, *Toward a Theory of Network Gatekeeping: A Framework for Exploring Information Control*, 59 Journal of The American Society For Information Science and Technology 1493, 1496 (2008).

[59] Id., at 1506.

[60] As Aaron Shaw contents, in an analysis of user-moderated platform Daily Kos, "decentralized gatekeeping consists of numerous, micro-level interactions between individuals engaged in a particular collective endeavor." Aaron Shaw, *Centralized and Decentralized Gatekeeping Online: Political Discourse and Mobilization on Daily Kos*, Pre-Publication Draft, at 12.

[61] In Wikipedia, it has been found that coordination through the use of communication tools is sometimes a better predictor of article quality than the total number of editors that work in a specific article. See Aniket Kittur and Robert E. Kraut, *Harnessing the Wisdom of Crowds in Wikipedia: Quality Through Coordination*, CSCW'08, November 8–12, 2008, San Diego, California, USA, 44 (2008).

[62] See A. Michael Froomkin, *Habermas@discourse.net: Toward a critical theory of cyberspace*, 116 Harvard Law Review 749 (2003), describing the adoption of Jürgen Habermas' rough consensus by the Internet Engineering Task Force in their decision-making processes.

common parameters for filtering serve as guidance for all users, not as coercive authority such as the rule of law[63].

There are certain aspects of how the platform is designed that facilitate user filtering. If the environment is shaped to allow for reputation monitoring, where the identity of users is clear, and certain modes of user surveillance by users themselves are built in, filtering can be more precise and effective[64]. Suppose a filter confirmation mechanism is established, whereby a post or file is only blocked once three different users decide it's illegal. When someone is considering if they should add the third "vote" for a block, trust on the user who made the first "vote" can influence their assessment. If that first user has had its block decisions confirmed in 95% of the cases, that third user can devote less work into evaluating whether or not to add the third filter order. Furthermore, these trust and collaboration mechanisms can be made to allow one user to profit from the viewing decisions of another user[65], such that content that is less and less viewed over time could be more vulnerable to censor "votes" than content that is widely shared and read.

Problems obviously arise from the reliance on user filtering. At least two can be identified upfront: incentives to engage in filtering and the tendency to over-filter. The prohibition on the exchange of child pornography material is perhaps the only worldwide consensus in the field of internet governance. If a company wants to give users the tools necessary for collective filtering of pedophilia on its platform, it need not worry whether or not users will employ them. The self-governance aspect of user-filtering would be largely eroded if users were offered money to perform this task. There's the risk of a backlash against the company[66]. The average internet user would actively engage in censoring instances of child pornography and would gladly denounce and exclude other users responsible for these violations, such that no monetary compensation is required. What is crucial here is that the filtering that a company needs to have accomplished on its platform is based on values that aren't always shared by the users. Fighting child

---

[63] Which is why decentralized gatekeeping isn't completely useless without codified, agreed-upon rules. Codification here serves a purpose of guidance, not legitimation: "even if users consent to being governed by community norms, they often have no idea what they are consenting to, and more important-ly, they have no ability to find out other than through trial and error. There are no pre-announced, publicly available, attainable, written, forward-looking, impartially enforced rules." Michael Risch, *Virtual Rule of Law*, 112 W. Va. L. Rev. 1, 35 (2009).

[64] This is especially true in online virtual worlds where MMOGs are played. "Reputation is a key element of social value to many players. The accumulation of social status is part of the reward for participation. Players institute their own regimes of surveillance." Sal Humphreys, *Ruling the Virtual World. Governance in Massively Multiplayer Online Games*, 11 European Journal of Cultural Studies 149, 162 (2008).

[65] "Using previous experiences from users who change options easily, it is possible to further expand the role of ratings in structuring large-scale online conversations to provide customized, worthwhile content to a heterogeneous community of users." Cliff Lampe et al., *Follow the Reader: Filtering Comments on Slashdot*, *in* Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI'07). San Jose, CA. April 28-May 3, 2007, 1253, 1261 (2007).

[66] The recent discovery of a kind of "rulebook for filtering" that paid censors receive from Facebook has caused some revolt by Facebook users. In this case most of the disappointment by users was directed at the rules themselves, not the widely known fact that Facebook outsources the job of filtering to poorly-compensated workers. However, this incident shows that the company is exposed to criticism precisely because of this practice, regardless of what the enforced filtering criteria are. *Inside Facebook's Outsourced Anti-Porn and Gore Brigade, Where 'Camel Toes' are More Offensive Than 'Crushed Heads'*. Available at: http://gawker.com/5885714/ (last visited Apr 26 2012).

pornography and hate speech usually are; banning the exchange of copyrighted works usually isn't[67]. As the presence of incentives are a fundamental element in a commons-based enterprise, it is likely that decentralized, user-performed filtering of copyright infringement wouldn't succeed. The second problem is excessive filtering: when given power, users have a tendency to gradually apply stricter standards and filter more and more content. Social networks are constantly troubled by this and Facebook recently had to face the online and offline wrath[68] of mothers who mobilized against the removal of pictures where women are shown breastfeeding[69]. This calls for mechanisms that operate as a check on the user filtering decisions. This restraint doesn't need to come from the direct intervention of the company in each case, overruling a user's decision to delete certain content. Other tools of checks and balances, such as distributed trust-building, multiple confirmation requirement and strict review and transparency of the actions by users with records of high number of filtering attempts all ensure the continuance of bottom-up, decentralized filtering.

V. Conclusion

Addressing the issue of internet intermediary liability is absolutely critical to the protection of an online environment that is conducive to innovation and that fosters freedom of expression. The developments in this legal field over the years have brought about the continuous risk for companies that operate online platforms such as social networks, peer-to-peer networks and virtual gaming worlds. The current legal environment in the United States, and especially for so many of the companies that wish to conduct business concomitantly in several countries, raises uncertainty about liability for the actions of users such that resigning to filter content isn't an option. Conversely, internet users are progressively adopting a very critical view of the censorship performed by these companies and are successfully organizing movements and isolated protests that push back against content filtering done by the platform provider.

In this setting, enabling and stimulating users of the platforms to filter illegal content themselves appears as an alternative that has great potential in building on top of the resilient objection to external, bottom-down control of the internet that netizens have asserted with great force on the early days of the web. User filtering is compatible with

---

[67] The notion that exchanging copyrighted content online is morally acceptable is especially prominent among teenagers, as studies have shown that they completely differentiate between material and immaterial theft (http://newsroom.unl.edu/releases/2011/05/03/Study%3A+To+college+students,+shoplifting+and+music+piracy+are+worlds+apart) and on average have 842 illegally downloaded songs on their portable devices (http://www.thetechherald.com/articles/Study-digital-music-piracy-is-rampant-amongst-teens/618/).

[68] Emil Protalinski, *Breastfeeding women protest outside Facebook offices*. Available at: http://www.zdnet.com/blog/facebook/breastfeeding-women-protest-outside-facebook-offices/8673 (last visited Apr 26 2012).

[69] The company itself was making the final decisions on picture removal, but it relied on user input to identify them, as they made clear in a press release: "It is important to note that any breastfeeding photos that are removed – whether inappropriately or in accordance with our policies – are only done so after being brought to our attention by other Facebook users who report them as violations and subsequently reviewed by Facebook." Emil Protalinski, *Facebook clarifies breastfeeding photo policy*. Available at: http://www.zdnet.com/blog/facebook/facebook-clarifies-breastfeeding-photo-policy/8791 (last visited Apr 26 2012).

the notion of self-government by internet users and might thus work in certain platforms where a sense of community has developed among the users or customers. This alternative solution requires mechanisms to be encoded into the company's platform that would enable users to red-flag content, view the reputation of each other on the platform and collectively coordinate guidelines for how the filtering would be exercised. Despite being a promising substitute for platform manager-controlled central filtering, user filtering suffers from problems like lack of incentives to censor certain content (especially that which infringes copyright) and the tendency to gradually over-filter.

User filtering has the potential to address the liability risk of online intermediaries, currently one of the main problems in cyberlaw. Unfortunately, there is a dearth of research on how decentralized gatekeeping could substitute for company-imposed content filtering, such that further study on this subject is required to better evaluate the possibilities for the success of user filtering in addressing the problem of online intermediary liability.