Let the Users be the Filter? Crowdsourced Filtering to Avoid Online Intermediary Liability[1]

Ivar A. Hartmann[2]

I. Introduction

Online platforms for decentralised 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 with one another, they play a crucial role as intermediaries. Because they are the managers of online communities where – just like in the real world –infringements of the law can occur, these companies are constantly sued by users or third parties alleging that they have a responsibility for what is done on their platform.

Even though safe harbour[3] provisions exist in American and EU law that release intermediaries from a duty to proactively monitor and filter user activity on their platforms, the liability standard is constantly shifting. Copyright owners' pleas, for example, demanding a different, less passive role for intermediaries have been gaining ground recently. The most prominent examples of this trend in recent times are, first, 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[4] and the "right to be forgotten" ruling by the Court of Justice of the European Union.[5] In the former, a bold challenge of the longstanding safe harbour in the Digital Millennium Copyright Act against strict liability for copyright violations was not summarily dismissed by the appeals court. In the latter, the court created a dangerous precedent

[1] A previous version of this paper was presented at the Oxford Internet Institute's The Internet, Policy & Politics Conference in 2014. See http://ipp.oii.ox.ac.uk/2014/programme-2014/track-b-policy/information-law-regulation-and-ethics/ivar-hartmann-let-the-users-be-the.

[2] Professor and researcher at the FGV Law School in Rio de Janeiro. MSc in Public Law (Catholic University of Porto Alegre). LL.M. (Harvard). Doctoral Candidate (State University of Rio de Janeiro).

[3] Safe harbour provisions, in this context, are legal rules that exempt online intermediaries from liability provided that they remove content deemed illegal upon a notice – by the offended party or by a court.

[4] *Viacom International, Inc. v YouTube, Inc.* (2013) No. 07 Civ. The case was later settled, which goes to show exactly how safe Google felt the safe harbour provisions to be. See Joe Silver, 'Viacom and Google settle $1 billion YouTube lawsuit' <http://arstechnica.com/tech-policy/2014/03/viacom-and-google-reach-settlement-in-long-running-youtube-lawsuit> accessed 18 August 2014.

[5] Google Spain SL, Google Inc. v Agencia Española de Protección de Datos (AEPD), Mario Costeja González (2012) Case C-131.

by classifying Google as a data controller instead of a content intermediary, thus creating a risk that any social network or forum be denied the safe harbour awarded to intermediaries.

These developments create an environment where safe harbour provisions no longer offer the same protection they previously did against strict 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 a 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 favour and against having users themselves filter content – 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 it simply would not exist.

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

For many years the idea that behaviour on the internet could not be regulated was very popular. It was a completely new world where the entities that exercised regulation either could not 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."[6] It is remarkable that this held to some extent true for a short while in the early days of the internet and then the "world wide web" – especially since the internet was born as a research project of the United States military. That may very well have been true while the internet was still in its academic and hippie era. During the 1990s, however, after the era of online academia and hippie dominance, the internet was taken over by the market logic, or "commodified."[7]

[6] John Perry Barlow, 'A Declaration of the Independence of Cyberspace' <https://projects.eff.org/~barlow/Declaration-Final.html> accessed 23 April 2012.
[7] This phenomenon has been described by many authors. See, for instance, Graham Murdock and Peter Golding's explanation: "Economically 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

Notwithstanding the appropriateness of understanding the internet as a new and different place,[8] 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 does not automatically make it an isolated place: web users are the same people who live within the borders of nation states and even those who do not access the internet are nonetheless affected by it. Total separation, although legally possible with the recognition of an independent cyberspace jurisdiction,[9] is unpractical and unreal. The contention that it is impossible to track information flow online was perhaps partially true up until the mid-1990s. However, the use of Deep Packet Inspection[10] and other mechanisms has allowed internet service providers (ISPs) and governments to constantly and effectively control online communication. Governments, in particular, seem to be intent in recent years to make up for lost time with "strenuous reassertions of national authority."[11] According to some accounts, in certain countries the government

political inequality' in Robin Mansell (ed), *The information society. v. III (Democracy, governance and regulation)* (2009), 15. 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* (2000), 308.

[8] "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' [1996] 48 Stanford Law Review 1403, 1403. See also Colin Crawford, '*Cyber*place: defining a right to Internet access through public accommodation law' [2003] 76 Temple Law Review 225.

[9] "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' [1995] 48 Stanford Law Review 1367, 1378. The authors do not, however, affirm that cyberspace and the physical world are perfectly separable.

[10] 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?' <http://www.pcworld.com/article/249137/what_is_deep_packet_inspection.html> accessed 26 April 2012.

[11] "States lay claim to geographic names and the representation of linguistic scripts in cyberspace; they scale up their surveillance capabilities; they make plans to weaponized cyberspace and 'secure' their part of it; they try to set themselves up as gatekeepers who can censor content." Milton L. Mueller, *Networks and States. The Global Politics of Internet Governance* (2010) 253.

controlled the internet from the very beginning, so that there never even was an initial golden period of freedom. [12]

Another common argument to support the unfeasibility of governmental regulation online 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.[13] 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.[14]

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 to adapt to the needs of private companies. For-profit websites covered the landscape and the cyberflâneur was gone.[15] 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 communication network the potential of which was lauded as revolutionary,[16] the fact remains that it is perfectly possible to regulate internet behaviour and this has been done for many years now.

A completely different issue is whether the internet *should* be regulated in the first place, especially by nation-states. Most of the current arguments for multistakeholderism in international internet governance[17] and self-regulation by

[12] See James Curran, 'Reinterpreting the Internet' in James Curran, Natalie Fenton, Des Freedman (eds) *Misunderstanding the Internet* (2012) 12.
[13] 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' [1996] 28 Conn L Rev 1095, 1126.
[14] "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' [1998] 65 University of Chicago Law Review 1199, 1234.
[15] "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 — has not 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' <http://www.nytimes.com/2012/02/05/opinion/sunday/the-death-of-the-cyberflaneur.html?pagewanted=all> accessed 23 April 2012.
[16] See Jonathan Zittrain, 'Be Careful What You Ask For: Reconciling a Global Internet and Local Law'. (Harvard Law School Public Law Research Paper No. 60, 2003) <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=395300>.
[17] 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

the private sector are built on top of beliefs shared by many authors who in the late 1990s and early 2000s openly rejected government regulation of the Net, even assuming that it could technically be done. Even today, authors call for an internet governance framework that emphasises the necessary diversity of the stakeholders,[18] avoiding a predominance of state power.

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,[19] 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.[20] A different proposed model was the recognition, by the nation state, of a new type of rulemaking process – one that is not 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, [21] accepting the decisive regulatory role played by choices on how the internet architecture is configured.[22]

---

and Global Governance: Principles and Norms for a New Regime' [2007] 13 Global Governance 237, 250: "[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 Robin Mansell (org), *The information society. v. III (Democracy, governance and regulation)* (2009), 384.

[18] See Luca Belli, 'A heterostakeholder cooperation for sustainable internet policymaking' [2015] 4 Internet Policy Review 2. For an example of this notion put into practice at a high-level internet governance event, Brazil's 2014 NETMundial, see the detailed description in Marilia Maciel, Nicolo Zingales, and Daniel Fink, 'The Global Multistakeholder Meeting on the Future of Internet Governance (NETmundial)' in *Multistakeholder as governance groups: observation from case-studies*. Research Publication 2015/001 (Berkman Center 2015).

[19] "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' [2000] 88 California Law Review 395, 417.

[20] 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' [1998] 5 Ind J Global Legal Stud 512, 535-539 and 542.

[21] "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' [1998] 76 Texas Law Review 553, 571.

[22] 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 behavior after it has been engaged on. Lawrence Lessig, *Code: And Other Laws of Cyberspace, Version 2.0* (2006) 7. 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

Unlike in the cyberpopulist model, *lex informatica* would be enforced by government, such that the latter would merely lose its rulemaking prerogative[23], and even then only on what concerns human action online.[24] Even those who accepted enforcement of traditional legal norms, especially regarding commerce on websites, argued for concessions. A company could not 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 recognise that companies often targeted a specific audience despite the fact that their website was viewable to anyone.[25]

It is very important to notice that these models fundamentally evoke self-government, just as advocates of the impossibility of regulating internet did. John Perry Barlow's 1996 Declaration of Independence of Cyberspace symbolised 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.[26]

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. Interestingly, this idea was getting stronger in a

School' [1998] 27 The Journal of Legal Studies 661, 666. More recent literature has further developed Lessig's concept, arguing that "arrangements of technical architecture are also arrangements of power" and presenting current examples in the worldwide internet governance. Laura DeNardis, 'Hidden Levers of Internet Control' [2012] 15 Information, Communication & Society 720, 721.

[23] That is because "[*I*]*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 21, at 577.

[24] 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. 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.

[25] "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.

[26] "[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 19, at 432.

time when the United States government pushed to control the world wide web via the domain name system. [27] The netizen self-governance rational 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 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 behaviour. 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 explicit, voluntarily created and codified, much like legal norms.[28] It is obvious that these two types of self-imposed rules have an intrinsic relationship such as that of law and morals[29] and therefore an attempt at a clear split would be both unwise and difficult.

For the purposes of this paper, I assume that nation states can and should regulate the web. At this point it should be clear that the early literature on whether the internet could be regulated and how is relevant here only to show that there was always an interest on the part of users to exercise choice on the configuration of the rules and power in their enforcement. The question of *whether nation states can in fact regulate* internet behaviour is irrelevant here not only because they have effectively been doing so for years,[30] but also – and more importantly – because it doesn't negate or decrease the interest of internet users in playing an active and direct role in such regulation. The state's regulatory capabilities do not have to be nullified in order for other stakeholders to play a part. Much to the contrary: in Europe and the United States, there is a trend of moving from online company self-regulation to co-regulation, where the government dispenses more attention to internet activity.[31] In both systems, however, companies play a part along with nation states.

For the same reason, with the literature on cyber-anarchism theory mentioned here I do not intend to argue that nation states *should not regulate* the internet. That argument is barely tangential to the core discussion of intermediary liability that is my focus in this paper. What matters is that these works show users do not want be passive subjects of regulation. I suggest they have actually been feeding an instinct of self-government.

[27] See Jack Goldsmith and Tim Wu, 'Who Controls the Internet? Illusions of a Borderless World' (2006) 46.
[28] I'm adopting the distinction between customary norms and legal (positive) norms made by Hans Kelsen, *Pure Theory of Law* (2nd ed 1978).
[29] As described by H. L. A. Hart, 'Positivism and the Separation of Law and Morals' [1958] 71 Harvard Law Review.
[30] Goldsmith and Wu, *supra* note 27.
[31] Ian Brown and Christopher T Marsden, Regulating Code. Good governance and better regulation in the information age (2013).

My argument is merely that while internet users today often recognise the force of traditional regulation, the codified, written rules that the user communities spontaneously create undeniably demonstrate the assertion of a self-governance prerogative. More than in other contexts, people in online communities feel they are entitled to some rule-setting powers. In the third part of the paper, I intend to show that under adequate circumstances, a private company could harness this enthusiasm.

In the case of speech, that means users making censoring or filtering decisions. We face a scenario where users themselves set rules on allowed and forbidden content, albeit under the auspices of traditional state regulation and with the cooperation of private companies. The first part of the paper has so far suggested that users wish to take on that task. The second part is about the convenience for companies that users do so. The third part discusses the technical mechanisms required to carry this out. It would be pertinent, however, to ponder on the legality of crowdsourced filtering.

Concerning the flow of information, the law has always preoccupied itself with tailoring the conditions in which speech might be censored for being abusive of a third party's individual rights. Historically, rule making on speech has thus focused on the details of excessive speech and how it should be constrained. There are many legal dispositions with limitations for speech – such as libel and copyright. That is because the path of least resistance was for expression to flow naturally. Under such traditional regulation of speech, the only concern with crowdsourced filtering would be to set the limit for speech that legitimates users in their task of taking down content. In short: law must define what of their own speech private parties should refrain from uttering.

However, one of the internet's many collateral effects has been a shift in the power of private parties to express themselves. The more noticeable and discussed aspect of such a shift is that there is no longer any scarcity of space and everyone can potentially be heard by everyone. The less discussed and perhaps more decisive aspect is that the platforms for speech are now owned and operated by private parties. When public spaces and the main forum, censorship rules concern state action. However, when private spaces – social network newsfeeds, search engine results – become the main forum, censorship rules concern private action. In short: law must now define what speech of others private parties should be forced to tolerate.

Private censorship is one of the most daunting problems of internet regulation right now. The law is severely late to address this and state action doctrine is one of the obstacles. In Germany the constitutional tradition establishes the duty also of private parties to respect constitutional rights. The German Constitutional Court

demonstrated in the Lüth case, half a century ago, that individuals can harm censor speech just as dangerously as the state and should therefore be constrained in their efforts to do so.[32]

If and when rules are set to define what kind of speech companies must allow in their platforms and their policy guidelines, these laws will also have to answer whether and to what extent the exercise of filtering by users is compatible with proper protection of the right to freedom of expression. It should be noted that this challenge is not related to separating speech that is allowed from forbidden speech. Rather, as the other major challenges of protecting free speech in the digital age, it is about the institutional design of platforms and the law itself, which includes choices on procedural remedies and who gets to make the decisions about content.[33] The risk of abuse when society moves decisions on free speech from the courts to the hands of private companies or users is not negligible. Where groups of users dictate what can and cannot be said, the rise of a tyranny of the majority is high – especially with our current legal tradition. That is the main legal challenge one could pose against crowdsourced filtering.

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 start-ups incorporate into 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.[34] 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.[35] 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. 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?

[32] BVerfGE 7, 198 I. Senate (1 BvR 400/51)
[33] Jack Balkin, 'Digital Speech and Democratic Culture: A Theory of Freedom of Expression for The Information Society' [2004] 79 1 NYU L Rev 1.
[34] See Jeff Jarvis, *What Would Google Do?* (2009).
[35] 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).

Should they? Will they be liable when users in their platform violate the privacy or property of third parties?

A brief summary[36] is in order before I lay out the specific cases in more detail. The rules that govern online intermediary liability have nuances in the United States and Europe depending on the type of intermediary and nature of the rights that were violated by the users. The more editorial and content-management power the company has, the stricter the standard.

Regardless of the liability standard set by statute or jurisprudence, rights holders are constantly pushing for an expansion in the filtering obligations of intermediaries. In this section of the paper, I intend to show that such a push has been at least partially successful in most jurisdictions, increasing the need of intermediaries to search for means of filtering.

In the mid-1990s, intermediary liability was noticed as an entirely new and incredibly relevant issue. Companies had never 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 were not 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 would not exist.[37] The solution found, however, was to give immunity to intermediaries when users exchange data that infringes copyright[38] or

[36] Good overviews of the scenario for intermediary liability before the changes I describe in this section are provided by Benoît Frydman and Isabelle Rorive, 'Regulating Internet Content through Intermediaries in Europe and the USA' [2002] 23 Zeitschrift für Rechtssoziologie 1 and Yulia A. Timofeeva, 'Hate Speech Online: Restricted or Protected? Comparison of Regulations in the United States and Germany' [2003] 12 Journal of Transnational Law & Policy.

[37] 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' <http://www.wired.com/threatlevel/2008/10/ten-years-later/> accessed 26 April 2012.

[38] 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

constitutes lewd speech.[39] In theory, this would have meant that internet companies were safe from a big headache and could further conduct their business unhampered by fear of liability. But the current reality is much different.

Firstly, the adoption of a safe harbour for intermediaries in the United States was not 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,[40] similarly to what had been done by American legislation. This has proved to be a not-so-safe harbour 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.[41] The solution found, however, was to give immunity to intermediaries when users exchange data that infringes copyright[42] or constitutes lewd speech.[43] In theory, this would have meant that internet companies were safe from a big headache and could further conduct their business unhampered by fear of liability. Nevertheless, the current reality is much different.

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

content."

[39] 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."

[40] Directive 2000/31/EC of the European Parliament and of the Council [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 [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 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."

[41] The video had been posted to the Google Video website in 2006. See Stephen Shankland, 'Execs convicted in Google Video case in Italy' <http://news.cnet.com/8301-30685_3-20000092-264.html?tag=newsEditorsPicksArea.0> accessed 24 April 2012. The executives were eventually acquitted, which does not entirely assuage concerns of online intermediaries. Jon Brodkin, 'Italy finally acquits Google execs convicted over user-uploaded video' <http://arstechnica.com/tech-policy/2012/12/italy-finally-acquits-google-execs-convicted-over-user-uploaded-video/> accessed 3 June 2015.

[42] Section 512 of the Digital Millennium Copyright Act [1998].

[43] Section 230 of the Communications Decency Act.

customers.[44] 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.[45] 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.

In both cases, the reasoning of the Court was that imposing an absolute blanket-censorship obligation on ISPs and social networks was a disproportionate 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 twelve years after the safe harbour 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 is yet another reason for online intermediaries to worry. Privacy protection was understood to trump safe harbour or at least call for a different, least protective interpretation of it. That is because the Court considered Google to be a data controller instead of a content intermediary. The reason was that Google conducted "organization and aggregation of information" producing a "structured overview of the information."[46] The fundamental mistake made by the Court was that it created a "right to oblivion" with the excuse of trying to protect a "right to erasure." The former relates to publications and expression, while the latter relates to personal data stored in databases (as opposed to published) and subject to automated processing (as opposed to readership.)[47]

The editing of third-party content that the ECJ found Google to be performing is precisely what Facebook's newsfeed is. Any intermediary that purports to offer its users a "structured overview" of the user content in its platform is in risk of being labelled a data controller. This is in rampant conflict with the e-commerce directive's

[44] Scarlet Extended SA v Société belge des auteurs, compositeurs et éditeurs SCRL (SABAM). [2010] Case C-70/10.
[45] Société belge des auteurs, compositeurs et éditeurs SCRL (SABAM) v Netlog NV [2010] Case C-360/10.
[46] "Also, the organization and aggregation of information published on the internet that are effected by search engines with the aim of facilitating their users' access to that information may, when users carry out their search on the basis of an individual's name, result in them obtaining through the list of results a structured overview of the information relating to that individual that can be found on the internet enabling them to establish a more or less detailed profile of the data subject.". C-131/12, *supra* note 5.
[47] Ambrose, Meg Leta and Jef Ausloos, 'The right to be forgotten across the pond' [2013] 3 Journal of Information Policy.

rules and creates precisely the level of risk and uncertainty[48] for the intermediary that the safe harbour rule was enacted to prevent. Any intermediary that purports to offer its users a "structured overview" of the user content in its platform is in risk of being labelled a data controller. Back when the right to be forgotten was introduced to the draft of the new EU personal data protection directive, legal scholars predicted that solution would have serious chilling effects.[49] The Court ruling turned out to be much worse. The ECJ found that Google was producing new information by organising and aggregating previously published information. Yet somehow the Court did not grant the company the safe harbour protection that the EU personal data protection directive guarantees to those who are in the business of disseminating information – such as the press.

Even legislators in Europe, who have been quarrelling with American tech giants in the past few years over tax evasion allegations,[50] came out against the ruling stating it is "unworkable."[51] That is 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."[52] The future for intermediaries in Europe is indeed grim.

The company has since been doing just that. Even if we disregard the fact that the ECJ ruling has pushed Google into exercising a role that was meant for courts – deciding what is and is not protected expression, the practical outcome is that the search engine is now vulnerable to litigation and liability if it disagrees with a user on whether specific search results should be taken down. That is precisely the opposite of what the safe harbour for intermediaries intended. Even those who do not oppose it in principle[53] acknowledge this uncertainty about the standard and application of the right to be forgotten as a problem.

---

[48] 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 5.
[49] Jeffrey Rosen, 'The right to be forgotten' [2012] 64 Stan L Rev Online 88.
[50] See Andrew Frye, 'Renzi Pressed to Put Google, Facebook Taxes on EU Agenda' <http://www.bloomberg.com/news/2014-07-01/renzi-pressed-to-put-google-facebook-taxes-on-eu-agenda.html> accessed 18 August 2014; Frances Robinson, 'France Pushes EU to Regulate U.S. Internet Companies' <http://online.wsj.com/news/articles/SB10001424127887324492604579085222987377040> 18 August 2014. See also House of Commons. Committee of Public Accounts, *Tax Avoidance–Google* (Ninth Report of Session 2013–14) <http://www.publications.parliament.uk/pa/cm201314/cmselect/cmpubacc/112/112.pdf> accessed 18 August 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."
[51] Alex Hern. 'Lords describe Right to be Forgotten as 'unworkable, unreasonable, and wrong'' <http://www.theguardian.com/technology/2014/jul/30/lords-right-to-be-forgotten-ruling-unworkable> accessed 18 August 2014.
[52] European Union Committee. Second Report, *EU Data Protection law: a 'right to be forgotten'?* <http://www.publications.parliament.uk/pa/ld201415/ldselect/ldeucom/40/4002.htm> accessed 18 August 2014, para 33.
[53] Rolf Weber 'The Right to Be Forgotten. More Than a Pandora's Box?' [2011] 2 JIPITEC 120.

The latest addition in the series of rulings eroding the e-commerce directive's safe harbour is *Delfi AS v. Estonia* (2015) ECtHR 64659/09. Delfi.ee, a major Estonian news outlet, published a story on ice bridges. The piece generated many reader comments, and some of them included threats to a certain individual. This person asked Delfi to remove the threats and pay damages for hosting them. The company removed the comments as requested, but refused to pay. A court later forced Delfi to pay damages, even though as an intermediary it had removed the allegedly abusive posts upon request. The European Court of Human Rights heard the case and decided that Delfi's liability did not violate the European Convention on Human Rights' free speech protection. This means that any intermediary that is sued for damages in Europe, even after having removed content upon request, cannot seek freedom of expression shelter under the Convention. The future for intermediaries in Europe is indeed grim.

The prospect in other countries is shaky at best. In 2014 Brazil enacted its *Marco Civil da Internet*, a groundbreaking landmark internet regulation statute[54] 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 harbour and tend to be solved by Brazilian courts with notice-and-takedown or something even worse. Moreover, the judiciary's track record is certainly a bad omen.

Provisions of the Brazilian 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. Brazil's Superior Court of Justice adopted this view for years, until it receded to a notice-and-takedown standard on a more recent ruling, which explicitly finds support on the merits of a similar case.[55] The Supreme Constitutional Court has picked up a similar case for judgment and could go one way or the other. In theory it is not even bound

[54] Glyn Moody, 'Brazil's 'Marco Civil' Internet Civil Rights Law Finally Passes, With Key Protections Largely Intact' <https://www.techdirt.com/articles/20140326/09012226690/brazils-marco-civil-internet-civil-rights-law-finally-passes-with-key-protections-largely-intact.shtml> accessed 18 August 2014

[55] 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' [2009] 70 Revista de Direito do Consumidor 41.

by the *Marco Civil* choice for court-order-and-takedown because the Justices could easily rule that the Constitution requires more effective protection for defamation victims.

The second reason why companies cannot completely ignore the content of exchanges in their platforms is that the safe harbour rule has caveats and the result of judicial interpretation over the last ten years has not been entirely favourable to intermediaries. In the United States, for example, 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 characterise editorial action and thus trigger liability. This has been the case for peer-to-peer software, where since Napster the developers have gradually decentralised control of the file exchange process, relying more and more on the sense of centralized coordination below a certain threshold over which indirect liability ensues.[56]

Two cases that reached federal appeals courts show that the distinction between a liable intermediary and one that is in safe harbour 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 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 acted as a content co-developer and whether it had induced infringement.[57] Both questions were answered in the affirmative for Roommates.com.

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

[57] "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.

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 did not 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.[58] However, Judge Easterbrook explicitly denied that safe harbour 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.

Grokster was not about a social network or another type of website, rather it concerned a peer-to-peer 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. The concept of platform is itself increasingly more fluid. Airbnb, for example, is used both as website and an app and it must be careful, in both contexts, not to commit the same error as Roommates.com.

Metro-Goldwyn-Mayer Studios Inc. v. Grokster, Ltd., 545 U.S. 913 (2005) marks a strong shift away from anything resembling a safe harbour 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,[59] in Grokster the Court ventured into the intentions of the platform developer. Regardless of the possibility of using the platform for legal purposes, the manager could be held liable if it had induced infringement.[60] Of course Justice Souter in Grokster did all he could to make it seem as though the Sony rule was not being abandoned, but the fact is the rule changed.[61] This made intermediary liability more uncertain than it was before the ruling.[62] This was only the culmination of an ongoing

[58] 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.

[59] "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.

[60] "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.

[61] "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* (2008) 77.

[62] "[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' [2006] 78 Swiss Review of Business Law and Financial Market Law 119, 125. This

process in lower courts, where the changes being made to the law were increasing the uncertainty for platform developers.[63]

Safe harbour for online intermediaries has thus been turned into an indirect liability standard, one that inevitably curtails legitimate use that individuals may make of a legitimate online platform.[64] Furthermore, advocates of the fight against online child pornography call for a revision of the safe harbour provision in the U.S. Communications Decency Act (CDA),[65] which shields companies from civil liability – except for intellectual property issues. Criminal liability, on the other hand, is being revamped by revenge porn legislation that is now appearing in more and more countries, requiring more caution from intermediaries than ever before. [66]

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

---

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." James Boyle, *supra* note 62, 79.

[63] 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' [2006] 19 Harv JL & Tech 253, 285.

[64] "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' [2003] 16 Harv JL & Tech 395, 409. 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' [1998] 74 NYU L Rev. 354, 393.

[65] 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 do not 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"* <http://arstechnica.com/tech-policy/news/2009/03/a-friendly-exchange-about-the-future-of-online-liability.ars> accessed 24 April 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' <http://ssrn.com/abstract=1559208> accessed 24 April 2012, 3.

[66] See Mary Anne Franks, 'Drafting An Effective "Revenge Porn" Law: A Guide for Legislators' <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2468823> accessed 3 June 2015.

This carries its own set of problems, of course. First, engaging in filtering has a cost. Companies commonly allude to the impracticability of exercising human oversight over the activity in online platforms. Automated filtering is not at all 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.[67] Liability poses a big financial threat, one that is not 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 in a difficult situation: if it filters content, consumers potentially react badly, organise and protest;[68] if it does not filter, it highly increases the chances of being held liable – and sometimes incurring millions of dollars in 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.*[69] 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.[70]

[67] This predisposition for censorship amounts to a serious chilling effect on free speech. The 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' [2010] 24 Harv JL & Tech 171. The author diagnosed the problem in a system of strong safe harbour, 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.

[68] 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* (2006) 220.

[69] Liz Miller, 'Google's Viacom Suit Legal Fees: $100 Million' <http://gigaom.com/2010/07/15/googles-viacom-suit-legal-fees-100-million/> accessed 18 August 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.

[70] 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' <http://www.nytimes.com/2012/01/28/technology/when-twitter-blocks-tweets-its-outrage.html?pagewanted=all> accessed 25 April 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.

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

The ideal of self-governance by internet users was very prominent in the late 1990s and early 2000s. However, momentum has been lost due to the realisation 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 direct or indirect norm-enforcing roles. This arises 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.

Two comments are warranted at this point. First, I am discussing an option that private companies might exercise to lower their costs of operation. They could employ internet users' motivation for self-governance as a means to an end. The platform owners would therefore retain last-word authority over code and architecture choices as well as over individual user decisions.[71] User filtering, in this sense, is not a first step for users to liberate themselves and take over. The latest developments at Reddit illustrate this very well: a forum that is highly renowned for the libertarian views of its users when it comes to freedom of expression is currently taking initial steps to curtail instances of serious bullying and harassment.[72]

Second, this paper's main proposition operates within the current Western legal paradigm, including the protection of free speech and property as individual rights. Nevertheless, it could be read as a first step to a much more profound change, one that would require an entirely different framework of peer-to-peer elements in the development of law itself.[73]

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.[74] This is

[71] That is my assumption in this article. For a more detailed discussion of why that is the case, especially from a Marxist perspective, see Christian Fuchs (2012) 'Critique of the political economy of web 2.0 surveillance' in *Internet and surveillance: the challenges of web 2.0 and social media* (Routledge, New York, 2012), 31-70.
[72] Julia Greenberg, 'Reddit wants to exile trolls. But growing up is hard' <http://www.wired.com/2015/05/reddit-wants-exile-trolls-growing-hard/> accessed 3 June 2015.
[73] See Melanie Dulong de Rosnay, 'Peer-to-Peer as a Design Principle for Law: Distribute The Law' [2015] 6 Journal of Peer Production. While I do not intend to support this bolder argument here, I believe it is compatible with the core ideas behind this paper.
[74] Studies find, for example, that posting as well a reading posts from other users have a positive impact on a participant's *sense of virtual community* (see Lisbeth Tonteri et al. 'Antecedents of an experienced sense of virtual community' [2011] 27 Computers in Human Behavior); that attachment of the individual to a platform based on identity with the group is easy to achieve and that such attachment varies according to the type of online community (see Yuqing Ren et al. 'Building Member Attachment in Online Communities: Applying Theories of Group Identity and Interpersonal Bonds' [2012] 36 MIS Quarterly 3); that "Most if not all of [Wikipedia's] growth is grassroots and bottom up. This growth is not explained by traditional vectors of funding, fiat, or momentum. Instead, the multidimensional, sociological, and psychological motivations of

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.[75] 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.[76] 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 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.

One successful example of crowdsourcing filtering of user-generated content is 9GAG.[77] Anyone can become a user and post something – usually an image. Posts vary in meaning and purpose, but the majority is humorous and revolves around common internet memes. A user's post goes to the voting page.[78] If it manages to attract support from enough users through a voting system, the post will then be shown in the first page. 9GAG users will usually spend most of their time in the first page, where they will not see the posts that never garnered enough popularity. Crowdsourced filtering on 9GAG is mostly positive filtering, but the result is nevertheless that some posts will not be visible to the average user due to

individual contributors take center stage." (Sheizaf Rafaeli and Yaron Ariel, 'Online Motivational Factors: Incentives for Participation and Contribution in Wikipedia' in A Barak (ed) *Psychological Aspects of Cyberspace* (2008)).

[75] 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 69, 258.

[76] 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 Jack Balkin and Beth Noveck (eds) *The State of Play. Law, Games, and Virtual Worlds* (2006) 24.

[77] http://www.9gag.com.

[78] http://www.9gag.com/fresh.

decentralized filtering conducted by other users. The managers of the website hardly have to worry about, for example, a child abuse post reaching the average users.

In the example of 9GAG, user action has direct effect on whether content is visible to others. It is not removed from the website altogether just for not being popular. Nevertheless, filtering doesn't only mean the identification and removal of content, it also means separation and categorisation of content. If a platform has managed to have users successfully categorise posts over time, then it is one important step closer to stimulating users to remove posts with certain abusive contents.

Even then, managing visibility or availability of content is not the only foreseeable role that users can play. Rather, it is one type within a large variety of options, ranging from more to less direct user influence on what content is displayed. Users could, for example, merely contribute with knowledge about what content is worthy and unworthy, making the software that actually filters fundamentally more accurate.[79] The more indirect the user input is, however, the more susceptible the intermediary becomes to the costs of filtering mentioned earlier.

There is a fine line between indirect user filtering and company filtering. The ideal solution for some companies might be an algorithm with machine-learning features that takes qualitative user input (e.g. not merely red flags) on what content should be taken down. One reason for that is if user input feeds human judgement calls instead of a machine-learning-enabled algorithm, then the company could be vulnerable to biases in its evaluation of the information. [80]

To illustrate the problem, suppose Facebook had one large team of employees located in the United States reviewing all of illegal content flags pointed out by users in different countries. Due to their personal biases, this group of people would very likely interpret reports of indecent pictures made by American users and those made by Iranian users in a significantly different fashion. Prejudice in the revision of user filtering could constitute cause for liability. The question then is: how accurate can an algorithm be at identifying content that should be taken down, given that it can draw help from user input? The job is highly complex, but advances have been made using the bias of users themselves to improve a software with a similarly complex task.[81]

[79] See Kuldeep Yadav et al., 'SMSAssassin: crowdsourcing driven mobile-based system for SMS spam filtering' [2011] HotMobile '11 Proceedings of the 12th Workshop on Mobile Computing Systems and Applications 1-6. The authors describe how user input helped calibrate an algorithm for filtering spam in SMS texts.
[80] See Henning Piezunka and Linus Dahlander, 'Distant Search, Narrow Attention: How Crowding Alters Organizations' Filtering of Suggestionsin Crowdsourcing' [2014] Academy of Management Journal, June. The study finds that when organizations crowdsource input they tend to favor suggestions more familiar to them over those that are distant.
[81] Pedro Henrique Calais Guerra et al., 'From bias to opinion: a transfer-learning approach to real-time sentiment analysis' [2011] KDD '11 Proceedings of the 17th ACM SIGKDD international

The main issue with crowdsourced filtering is whether users could successfully entertain a task of governance. Filtering is a form of censorship and would require, in some cases, shunning the users who engage in wrongdoing. It is useful to frame this as the decentralised, commons-based production of an information service: filtering content. This model of production depends on modularity, granularity and heterogeneity.[82] 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,[83] go a long way in providing for modularity and granularity. Heterogeneity seems to characterise the task as well: while certain content is more obviously infringing than others are, there are also those grey-area instances. There is the sale of copyrighted music and then there is 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.[84] At the same time, this is a peculiar kind of gatekeeping because it involves the traditionally gated becoming the gatekeepers,[85] the decision-makers on the issue of whether or not certain content passes scrutiny and can be shared in a community. This is decentralised gatekeeping whereby the users of an online platform purport to collectively fulfil a goal related to the control of information flow.[86] In order for user filtering to work, coordination is not as essential as in other collective endeavours like the production of encyclopaedia articles in Wikipedia.[87] 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

conference on Knowledge discovery and data mining, 150-158. ACM New York, NY, US.

[82] Yochai Benkler, 'Coase's Penguin, or, Linux and The Nature of the Firm' [2002] 112 Yale LJ 369, 435-436.

[83] 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." <http://boston.craigslist.org/cgi-bin/personals.cgi?category=stp> accessed 26 April 2012.

[84] 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' [2008] 59 Journal of The American Society For Information Science and Technology 1493, 1496.

[85] Id., at 1506.

[86] As Aaron Shaw contends, in an analysis of user-moderated platform Daily Kos, "decentralized gatekeeping consists of numerous, microlevel interactions between individuals engaged in a particular collective endeavor." Aaron Shaw, 'Centralized and Decentralized Gatekeeping in an Open Online Collective' [2012] Politics & Society 40, 357.

[87] 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' [2008] CSCW'08, November 8–12, 2008, San Diego, California, USA, 44.

insufficient censorship. This does not mean that without unanimity on a general set of rules the whole enterprise is doomed to fail. Rough consensus can play an important part in the decision-making process of online communities,[88] but in any event, the existence of some common parameters for filtering serve as guidance for all users, not as coercive authority such as the rule of law.[89] Crowdsourced work – free or paid – could benefit from some hierarchy mechanisms[90] in order to begin tackling the minimal coordination issue which, in this case, consists e.g. of avoiding overfiltering. Decentralised management of Wikipedia, for example, relies on different user strata in order to avoid too much or too little redaction of user edits.[91]

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.[92] Suppose a filter confirmation mechanism is established, whereby a post or file is only blocked once three different users decide it is illegal. When someone is considering if she should add the third "vote" for a block, trust on the user who made the first "vote" can influence her 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,[93] 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.

[88] See A. Michael Froomkin, 'Habermas@discourse.net: Toward a critical theory of cyberspace' [2003] 116 Harvard Law Review 749, describing the adoption of Jürgen Habermas' rough consensus by the Internet Engineering Task Force in their decision-making processes.
[89] Which is why decentralized gatekeeping is not 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' [2009] 112 W Va L Rev 1, 35.
[90] A better framework for the use of hierarchy mechanisms is an important research agenda for both free and paid online crowdsourced work, as pointed out by Aniket Kittur et al., 'The Future of Crowd Work' [2013] Proceedings of the CSCW'13, February 23–27, 2013, San Antonio, Texas, USA.
[91] For empirical evidence that provides a detailed account, see Dariusz Jemielniak, *Common Knowledge?: An Ethnography of Wikipedia* (2014).
[92] 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' [2008] 11 European Journal of Cultural Studies 149, 162.
[93] "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' [2007] Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI'07). San Jose, CA. April 28-May 3, 2007, 1253, 1261.

Problems obviously arise from reliance on user filtering. At least three can be identified upfront: incentives to engage in filtering, the tendency to over-filter and the skewed demographics of the users who engage in filtering.

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 paedophilia 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.[94] 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 are not always shared by the users. Fighting child pornography and hate speech usually are; banning the exchange of copyrighted works usually is not.[95] In the already mentioned case of Reddit, a platform that relies heavily on user self-governance, the users share an extremely expansive notion of free speech. Because this evidently hinders attempts to contain serious harassment, the company's new measures to discipline trolls seems to be to "change the social norms and values of the site, to create an emerging culture of free expression online that is sensitive to harassment."[96] It is decentralised, but it remains gatekeeping. The intermediaries still hold the reigns and could potentially influence user's values and practices. Naturally, it all depends whether companies can find and exercise the proper incentives. Companies would have a hard time shaping user motivation and purpose vis-à-vis an online platform in order to get libertarians to protect copyright or liberals to combat vicious harassment, but research suggests that might be possible. [97]

[94] 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'' <http://gawker.com/5885714/> accessed 26 April 2012.

[95] 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 ('Study: To college students, shoplifting and music piracy are worlds apart' <http://newsroom.unl.edu/releases/2011/05/03/Study%3A+To+college+students,+shoplifting+and+music+piracy+are+worlds+apart> accessed 15 October 2017) and on average have 842 illegally downloaded songs on their portable devices (Stevie Smith, 'Study: digital music piracy is rampant amongst teens' <http://www.thetechherald.com/articles/Study-digital-music-piracy-is-rampant-amongst-teens/618/> accessed 15 October 2017).

[96] Greenberg, *supra* note 73.

[97] That is because "users may continue to participate in a site for different reasons than those that led them to the site", according to evidence from Cliff Lampe et al., 'Motivations to Participate in Online Communities' [2010] CHI '10 Proceedings of the SIGCHI Conference on

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[98] of mothers who mobilised against the removal of pictures in which women are shown breastfeeding.[99] This calls for mechanisms that operate as a check on the user filtering decisions. This restraint does not 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, decentralised filtering.

The third problem that affects the chances of user filtering becoming a reliable and effective mechanism has to do with the digital divide that plagues most countries and negatively affects the diversity of online communities. Naturally, the lack of means to access the web and digital illiteracy pose a challenge that is not limited to user filtering. Graham, Straumann and Hogan have shown how "Wikipedia is characterized by highly uneven geographies of participation."[100] They find that users from developing countries focus on editing entries concerning developed countries, which suggests that a well-functioning filtering community on pages and groups related to a poor country might be difficult even when the first steps to overcome the problem of access have already been successful. The lack of diversity and the replication of offline predictors of engagement in governance (mainly gender) is a problem of online e-democracy schemes.[101] As with social inequality, digital illiteracy can only be fought gradually, with the implementation of digital inclusion policies. Above all, it is paramount that governments invest in broadband infrastructure as well as digital literacy, acknowledging that internet access is an autonomous constitutional right with a complex positive dimension.[102]

Human Factors in Computing Systems, 10-15 April, 2010, Atlanta, United States.

[98] Emil Protalinski, 'Breastfeeding women protest outside Facebook offices' <http://www.zdnet.com/blog/facebook/breastfeeding-women-protest-outside-facebook-offices/8673> accessed 26 April 2012.

[99] 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' <http://www.zdnet.com/blog/facebook/facebook-clarifies-breastfeeding-photo-policy/8791> accessed 26 April 2012.

[100] Mark Graham, Ralph K. Straumann and Bernie Hogan, 'Digital Divisions of Labor and Informational Magnetism: Mapping Participation in Wikipedia' [2015] Annals of the Association of American Geographers.

[101] Ralf Lindner and Ulrich Riehm, 'Broadening Participation Through E-Petitions? An Empirical Study of Petitions to the German Parliament' [2011] Policy & Internet: Vol. 3: Issue 1, Article 4. María Rosalía Vicente and Amparo Novo, 'An empirical analysis of e-participation. The role of social networks and e-government over citizens' online engagement' [2014] 31 Government Information Quarterly, 379–387.

[102] Ivar A. Hartmann, 'A Right to Free Internet? On Internet Access and Social Rights' [2013] 13 J High Tech L 297.

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 file-sharing networks and virtual gaming worlds. The current legal environment in the United States and Europe raises uncertainty about liability for the actions of users such that failing to filter content is not an option. The law consistently shifts with new rulings from higher courts revising or reinterpreting safe harbour protections. The push from copyright holders and ordinary people worried about their reputation is a strong deterrent to speech-protective standards for intermediaries. Even where the law is seemingly clear, legal costs from constant court battles are high enough to suggest that companies should think of filtering options.

This uncertainty is even more worrisome for companies that wish to conduct business concomitantly in several countries. The latest legal developments in Brazil were used as an example to illustrate a legal environment that repeats itself in many other Latin American, African and Asian countries – where intermediary liability standards are sometimes even harsher as a result of reduced free speech protection. This scenario might make one conclude that heavy filtering is a near win-win option for platform providers. That, however, is not the case.

Society – and internet users themselves, in particular – are progressively adopting a very critical view of the censorship performed by these companies. People are successfully organising movements and isolated protests that push back against content filtering done by the platform provider. This happens in a context where big platforms are already facing criticism for multiple reasons. American companies take fire at home for slacking off in the protection of user personal data and abroad for tax evasion. They also often suffer trying to balance their global accepted content rulebook, on one side, and local values and customs, on the other. Finding the sweet spot between too much and too little filtering is an extremely delicate and complex task that companies have to perform in a time when even small mishaps can have enormous legal and public relations costs. This is what I tried to show in the second part of this paper.

Enter the online community. 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. In the first part of this paper, I intended to show that this sense of community is as old as the Net itself and runs deep. I would argue that even casual

users of humour websites would be willing to play a part and dedicate a small portion of their time to help separate acceptable from abusive content. The empirical studies find that the instinct is there, the organisation mechanisms exist and success cases are plenty.

The point is that user filtering is compatible with the notion of self-government by internet users and might work in certain platforms where a sense of community has developed among the participants. This alternative solution requires mechanisms to be encoded into the company's platform. The third part of the paper presented the contribution of the literature on decentralised gatekeeping and crowdsourcing of platform management. The phenomenon is neither new nor temporary in social networks – both online and offline. Studies have produced evidence of what works and what does not, the types of interactions between users and the limitations of this model. Practical examples from the daily use of the internet attest that companies are gradually enabling 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. Decentralised or crowdsourced filtering is not an anarchist alternative to total control in the hands of the company owning the platform. The examples discussed reveal an interaction between the free action of users and light management mechanisms put in place and operated by the company.

User hierarchy, up and down-voting and the revision of user take down decisions by users themselves are all schemes that coexist and interact with the company's policy as well as the community standards for what posts are allowed. In a well-crafted platform, the company can afford to play a supporting role in the filtering process, while also reserving some swaying capabilities to eventually correct course if the community is overfiltering or being too shy about taking down certain types of content. After all, despite being a promising substitute for platform manager-controlled central filtering, it would appear 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 decentralised 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.