Human rights by design: The responsibilities of social media platforms to address gender-based violence online: gender-based violence online

Nic Suzor
Molly Dragiewicz
Bridget Harris
Rosalie Gillett
Jean Burgess
TessVan Geelen

Abstract:

Gender-based violence online is rampant, ranging from harassment of women who are public figures on social media to stalking intimate partners using purpose-built apps. This is not an issue that can be addressed by individual states alone, nor can it be addressed satisfactorily through legal means. The normalization of misogyny and abuse online both reflects and reinforces systemic inequalities. Addressing gender-based violence online will require the intervention of the technology companies that govern the commercial Internet to prevent and combat abuse across networks and services. We argue that international human rights instruments provide an opportunity to identify with more precision the responsibilities of telecommunications companies and digital media platforms to mitigate harm perpetrated through their networks, and ensure that the systems they create do not reproduce gendered inequality. Finally, we present initial recommendations for platforms to promote human rights and fulfill their responsibilities under the United Nations Guiding Principles on Business and Human Rights.

KEY WORDS: human rights, intermediaries, gender-based violence, platforms, governance, social media

I. The internet has an abuse problem

The internet has an abuse problem. Indeed, in late 2016, the editorial staff at *WIRED Magazine* wrote an open letter to the internet that explicitly acknowledged how gender-based abuse has exploded in tech culture. *WIRED's* editorial juxtaposed the utopian visions many had for the internet at its birth twenty-five years ago with the hostile environment today:

> You were supposed to be the blossoming of a million voices. We were all going to democratize access to information together. But some of your users have taken that freedom as a license to victimize others. This is not fine. … As you got bigger and stronger, more people wanted to talk—but some of them were jerks, or worse. Remember flame wars? You had no immune system, and you started to rot. Now that rot has turned to blight. And here we are (*Wired* staff, 2016).

A major part of the problem, as the *WIRED* editorial indicates, is the mistaken ideology of value-neutral technological development, which results in a dual failure to design for pro-social uses and guard against anti-social ones. As danah boyd (2015) points out, 'We didn't architect for prejudice, but we didn't design systems to combat it either'. It is in this context that the *WIRED* editorial reflects an increasingly common (yet still controversial) critique that internet engineers, platforms, and services have an obligation to protect users from abuse by private individuals or collectives of individuals.

In this article, we argue that human rights offer potentially valuable yet underutilised resources to compel action to ameliorate online gender-based violence (GBV), which we understand as GBV that is mediated, practiced, or amplified via

internet-based digital media platforms. In particular, the United Nations Guiding Principles on Business and Human Rights (UNGPs) provide a foundation for action by the internet intermediaries -- telecommunications providers, search engines, content hosts, digital media platforms -- that play a major role in mediating communication between internet users (OECD, 2010). Although they are not state actors, the decisions these intermediaries make in governing their networks have a direct impact on the everyday lives of their users (DeNardis and Hackl, 2015; Gillespie, 2017). Due to their central role in internet communication and the degree of their influence on contemporary social life, internet intermediaries have ethical, social, and human rights responsibilities to their users and others affected by their networks (MacKinnon et al., 2014; Belli and Zingales, 2017; Council of Europe, 2017).

Telecommunications corporations and digital media platforms have been slow to respond to online abuse and harassment, persistently representing themselves as neutral intermediaries. Indeed, there is little consensus about what actions they should take to mitigate harm perpetrated through their networks. Generally, and particularly in the US where many major internet companies are based, internet intermediaries are not legally responsible when their networks are used in abusive ways. In the absence of compelling legal obligations, "the roles and responsibilities of intermediaries in relation to online gender-based violence remain unclear" (Pavan, 2017, p. 63). In Part II, we show how the UN *2030 Agenda for Sustainable Development* explicitly links the elimination of gender-based violence and abuse to women and girls' human rights. We argue that multiple forms of gender-based violence and harassment are linked by cultural norms and structural inequalities (Kelly, 1988). Addressing online GBV, then, requires the active and careful participation of state and non-state actors involved in internet governance.

In Part III, we argue that human rights tools can be used as part efforts to encourage digital media platforms to address GBV on their networks. While the primary obligations to protect human rights rest with nation states, the UNGPs make clear that private entities including major national and transnational telecommunications firms and digital media platforms have human rights responsibilities. We suggest that the language of human rights provides leverage to persuade internet intermediaries to tackle GBV. However, like all rights-based approaches to non-state action, the utility of human rights tools largely depends on informal pressure and cultural change. While the ongoing development of a treaty on business and human rights presents one option to create binding obligations, states already have clear obligations to ensure that their domestic laws effectively require that private actors, including internet intermediaries, take action to address GBV online.

We conclude with a set of issues for state and non-state actors to address in order to ameliorate the harms of GBV online. This is not a problem that can be easily addressed solely by the further criminalisation of harmful acts perpetrated by individual users. Particularly for normalised abusive acts that are harmful in aggregate but often do not individually rise to the level of criminal behaviour, addressing abuse online requires the involvement of private actors in setting and enforcing acceptable non-legal standards. States have a major role to play here; there is an opportunity to begin work to develop laws and regulatory processes that encourage (or require) telecommunications and internet businesses to be accountable for addressing gender-based violence in a way that is compatible with human rights. We argue that we are only at the beginning of necessary global conversations about how to address gender inequality online and off. However, there are steps businesses can take now,

including (a) enhanced transparency about their impact on human rights and their internal governance processes; (b) an ongoing commitment to ensure that human rights concerns are identified, considered, and addressed in the design and deployment of technological tools; and (c) the development of clear and consistently enforced policies and procedures to respond to abuse on their networks.

II Gender-based abuse and human rights

Interest in online abuse and harassment as a human rights issue is growing. In her 2016 report, Šimonović, the United Nations Special Rapporteur on violence against women, outlined a new emphasis on 'online violence against women':

While the use of information and communications technology has contributed to the empowerment of women and girls, its use has also generated online violence. In the view of the Special Rapporteur, there is a need to examine this recent phenomenon, and the applicability of national laws to it, and to make recommendations for States and non-State actors to fight online violence against women and girls while respecting freedom of expression and the prohibition of incitement to violence and hatred, in accordance with article 20 of the International Covenant on Civil and Political Rights. (Šimonović 2016, § G.70)

This framing signals an impetus to substantive work to advance efforts to address GBV online as a human rights issue while respecting competing rights and responsibilities. This agenda specifically seeks to examine the obligations of states in prosecuting criminal forms of online abuse as well as the role of internet intermediaries in identifying, reporting, and rectifying harassment and abuse against women and girls (OHCHR, 2017). In other words, in order to secure the full enjoyment

of human rights, online GBV must be systematically addressed by state and non-state actors. Responses to GBV will need to target pervasive gender-based harassment, such as organised gender-based harassment campaigns targeting women who are public figures, as well as actions that are criminalised, like direct and explicit threats of violence. This is a key point shared by most human rights activity: invoking human rights obligations to address GBV online is not a call to criminalisation. Indeed, the 2012 *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression* cautions against government overreach, noting that legal means alone cannot address the underlying problem of hatred and that criminal remedies may have unintended consequences (La Rue, 2012).

Gender-based violence and abuse are causes and effects of inequality, presenting a barrier to social and economic development. There are economic as well as ethical reasons to address these harms. For example, the UN's 2030 *Agenda for Sustainable Development* includes the first stand-alone global development goal to 'Achieve gender equality and empower all women and girls' (Goal 5) (United Nations, 2015), which includes Target 5.1 'End all forms of discrimination against all women and girls everywhere' and Target 5.2 'Eliminate all forms of violence against all women and girls in the public and private sphere', explicitly linking gender-based violence and gender inequality.

Gender-based abuse online must be understood in the context of the broader landscape of gender inequality (DeKeseredy, Dragiewicz and Schwartz, 2017). Kelly (1988) proposed that multiple forms of abuse and physical violence against women comprise a continuum, where 'typical' and 'aberrant' male behaviours shade into one another. In other words, the continuum of sexual violence ranges from behaviours

which are normalised, such as sexual harassment in public spaces, to those widely recognised as criminal, like physical assaults. Abuses that women are subjected to at home, work, and leisure can create 'climates of unsafety' (Stanko, 1990) which threaten their rights and freedoms. Similarly, online harassment and abuse directly threaten the autonomy, dignity, and privacy of targets of abuse and constrain their ability to participate in contemporary social life (Citron, 2014). We argue that online and offline contexts are essential to women's contemporary human rights.

Key stakeholders have recognised the importance of online abuse for women. For example, the Association for Progressive Communications (APC) (2017) has termed the many forms that harm can assume 'technology-related violence against women'. The Internet Governance Forum (2015) calls it 'online abuse and gender-based violence against women'. The UN Broadband Commission for Digital Development Working Group on Broadband and Gender (2015) has addressed 'cyber-violence against women and girls'. These categories include commonly minimised or accepted behaviours which spark little response as well as those that are more readily recognised, problematised, and regulated (Citron, 2009; Citron, 2014; Henry and Powell, 2016).

For a long time, concerns about abuse online have come into conflict with free speech norms. Too often, freedom of speech has been thought to be directly opposed to measures to prevent abuse, and values of content neutrality are deeply entrenched in technical internet ideology. As a result, in technical discussions about internet architecture and in business decisions of major platforms, concerns about freedom of speech and apparent neutrality have hindered meaningful progress on interventions to reduce abuse. The over-emphasis on freedom of speech in the context of social media platform governance is, in part, a historical artefact of the US-

centricity of the industry, including Silicon Valley's origins at the juncture of technology and the counterculture (Turner, 2006), especially the particularly US version of heroic romantic individualism that shaped early utopian ideas about what the internet was for (Streeter, 2011). The discourse of freedom of speech is disproportionately dominant in debates about the governance of online environments, as it has been in US communications policy more generally (Napoli, 2001). Contemporary social media platforms have traditionally seen themselves heroic defenders of individual freedom against authoritarian states. Because of this, the fierce defense of free speech can overshadow or delegitimise claims of harms caused by online harassment and abuse (Croeser, 2016, Jane 2017). Online harassment, for example, is often excused as harmless 'trolling' (Turton-Turner, 2013). There is also a persistent idea that online speech is immaterial which can be traced to the early days of the internet, where 'cyberspace' was imagined as a virtual space of borderless freedom, and its potential to contribute to libertarian ideals was emphasised (Turner, 2006; Cohen, 2012). But this utopian vision of a borderless and disembodied cyberspace has always obscured the internet's very real capability to embed and amplify discrimination and abuse (Franks, 2011) - after all, the infamous 'rape in cyberspace', which occurred in the early 1990s, was enacted verbally, within a text-based multiplayer game (Dibbell, 1994).

The conflict between protecting freedom of speech online and preventing abuse is a false dichotomy that rests on a refusal to account for power among individual users: systemic discrimination and abuse have serious negative impacts on the agency and participation of people who experience them. It has long been recognised in international human rights discourse that systemic discrimination and abuse can limit the ability of women to seek, receive, impart, and use information and

participate fully in society (Hussain, 1997, 1998). Women who are directly impacted by online GBV may self-censor and discontinue their participation in online spaces (Amnesty International, 2018; Jane, 2017). At the same time, women who see online GBV, or perceive a threat of it, may also suppress their voices in this context (Lenhart et al., 2016).

It is important to recognise that while the internet can be a vehicle for abuse, it is also a channel for women experiencing GBV to seek support and access resources. More broadly, as digital media and communications technologies permeate everyday life, access to these technologies is important for women's social and political participation and access to life-enhancing services, healthcare, education, and economic opportunities (DeKeseredy, Dragiewicz and Schwartz, 2017). This aligns with the articulation of the Sustainable Development Goal target 5b to "enhance the use of enabling technology, in particular information and communications technology, to promote the empowerment of women".

Human rights discourses provide potentially powerful but as yet underutilised tools for eliciting discussion about and identifying strategies to combat online GBV. Human rights frameworks provide a foundation for basic rights and protections that is an important counterpoint to the false dichotomies between freedom of speech and protection from abuse promoted by libertarian logics. Human rights prohibitions against discrimination, hate speech, and incitement to violence and genocide violence make clear that rights to freedom of expression and conduct carry associated responsibilities not to impinge on the rights of others (Fyfe, 2017). These are challenges that human rights discourse and law have had to deal with, and there is a large body of existing knowledge that emphasises context-specific analysis and proportionate responses to resolve real conflicts between rights where necessary.

While there is no easy answer, human rights principles at least provide a shared vocabulary and approach that can be used to work through these challenges.

Addressing online GBV necessarily means dealing with a messy tangle of conflicting and interconnected rights, within a complex system occupied by many different globally dispersed actors. It is not something that can be addressed by states alone through the criminalisation and enforcement of the most widely recognised abusive behaviours. Addressing the "ceaseless flickering hum of low-level emotional violence" (Haque, 2016) familiar on the internet today is a key part of addressing gender inequality and ensuring that women have the ability to exercise their rights -- including freedom of speech. Importantly, however, Kelly's (1988) point in articulating the continuum of sexual violence was not to encourage the criminalisation of a broader array of abusive behaviours despite their aggregate harmfulness. Rather, she argued for recognition of and discussion about the normative and competing values and beliefs underpinning the range of abusive behaviours that are not recognised as crimes. While some new laws have been useful to some survivors of GBV online and off, criminal legal responses remain inconsistently applied, often along intersectional lines of structural inequality that continue to exempt wealthy, White men from consequences while posing disproportionate unintended consequences for poor women of colour. Accordingly, many criminologists and anti-violence advocates are challenging the idea that the benefits of criminal legal approaches to GBV outweigh the harms of mass incarceration, calling for more attention to confronting and changing the social norms and structural inequalities that contribute to GBV (Potter, 2015; Richie, 2012).

Achieving real change will require the active participation of a wide range of individual and organisational actors to create an online environment where people

are free to exercise their rights. Since international human rights law has historically struggled to clearly articulate how non-state actors should be expected to act to promote human rights -- and how these expectations can realistically be enforced, it is important to recognise the need for multifaceted action, from the macro to the micro level, to promote environments where rights can be exercised. In the next Part, we examine one part of this puzzle, the increasing use of human rights discourse to frame the responsibilities of online intermediaries to combat GBV.

III The responsibilities of telecommunications companies and digital media platforms

Historically, human rights law has focused primarily on the obligations of state actors. The traditional public/private divide enshrined in early international human rights regimes limited human rights responsibilities to states (Baderin and Ssenyonjo, 2016; McQuigg, 2011). As a result, the role private actors should play in promoting human rights remains an unsettled issue (Hasselbacher, 2010). One attempt to articulate these responsibilities is The United Nations Guiding Principles on Business and Human Rights (UNGPs), which outlines respective human rights obligations of states and the responsibilities of businesses (United Nations Human Rights Council, 2011). The principles are the first articulation of UN expectations for how businesses should respect human rights (Mares, 2012). They are a non-binding articulation of expectations that moves away from earlier attempts to impose actual duties on firms under human rights law (Bilchitz and Deva, 2013; Nolan, 2013).

The core of the Guiding Principles is that businesses have responsibilities to 'respect' human rights, which 'means that they should avoid infringing on the human

rights of others and should address adverse human rights impacts with which they are involved' (United Nations Human Rights Council, 2011, 13). This language of responsibility is becoming more common in internet governance debates (Laidlaw, 2017). For instance, Jørgensen and Pedersen (2017, p. 193) argue that internet intermediaries have an extended 'sphere of influence', that is, they have a responsibility to their immediate stakeholders, but also internet users who may be directly or indirectly affected by their practices. In recent years, reports from international human rights institutions have started to explicitly call on internet intermediaries to do more to to combat GBV online (see, e.g. Council of Europe, 2017; MacKinnon et al., 2014; Kaye, 2016). A recent joint declaration between the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression and the Special Rapporteur on Violence against Women calls for systematic investigation of online GBV and remedies that maintain robust protection of rights to privacy and expression (OHCHR, 2017).

Although the non-legal responsibilities of internet intermediaries may be ambiguous, "[c]ivil society initiatives have been crucial in unveiling intermediaries from their overall impunity and exposing their practices in relation to digital harms to public scrutiny" (Pavan, 2017, p. 72). Take Back the Tech!'s campaign "#whatareyoudoingaboutVAW", for example, pressured internet intermediaries like Facebook, Twitter and YouTube to articulate what they were doing to combat violence against women on their platforms (Baker, 2014). Meanwhile, new institutions and projects are developing to hold internet intermediaries to account for their impact on human rights. The UN Internet Governance Forum's Dynamic Coalition for Platform Responsibility has been working to explore what responsibilities private entities may have under human rights instruments, and how these responsibilities ought to be

monitored and enforced (Belli and Zingales, 2017). The Ranking Digital Rights (2017) project provides a detailed analysis of how well the explicit policies of telecommunications and internet firms perform against human rights standards, focusing particularly on the rights of privacy and freedom of expression. In a similar vein, Terms of Service and Human Rights project has examined the contractual rules of digital media platforms against human rights standards (Venturini et al., 2016). The APC commissioned report (Athar, 2015) gathered evidence to increase understanding of technology-related violence against women (VAW), and provided recommendations to fight technology-related VAW. Most recently, Amnesty International (2018) has reported on violence and abuse against women on Twitter, arguing that the company is failing to adequately meet its human rights responsibilities in this context.

Non-binding social pressure can sometimes be effective, and the Guiding Principles do have rhetorical power. The major social media platforms, for example, have developed advisory networks with civil society safety advocates in response to escalating concerns about online abuse and harassment -- although the ultimate effectiveness of these arrangements is not yet clear (Dragiewicz et al., 2018). Some internet intermediaries have taken self-regulatory steps to develop new de facto standards (potentially in order to avoid more direct regulatory action) -- as in, for example, the relatively quick uptake of a complaints mechanism for image-based abuse by search engines and social media platforms in the wake of several high profile scandals (Suzor, Seignior, Singleton, 2017). They have also taken other voluntary actions to reduce harm, for example, Chatterjee and colleagues (2018, p. 2) reported the findings of their research on apps advertised for and used in intimate partner abuse to Google which "took action against some of the apps that violated

Play Store policies due to a lack of persistent notifications or promotion of spyware or stealth tracking. In addition, Google is expanding their restrictions on advertisement serving for IPV-related queries".

These civil society attempts to hold technology intermediaries to account are promising, but only one part of the picture. Businesses face real, practical incentives that often conflict with their ability or willingness to respect human rights (Bittle and Snider, 2013; McConnell, 2017). Non-binding tools like the UNGP are not likely to be sufficient, on their own, to achieve lasting and fundamental social change and adequate protection for human rights. In order for states to discharge their duties to protect, they will need to actively monitor the extent to which human rights violations are being adequately dealt with through the voluntary actions of businesses. It is highly likely that states will need to invest in more direct regulatory and co-regulatory work in order to hold internet intermediaries accountable in their role in mitigating and responding to online gender-based violence. As a first step, states could immediately do more to explore how the guidelines relate to existing legal obligations under domestic law. In the longer term, ongoing efforts to develop a binding treaty on business and human rights may provide an avenue to better articulate the legal duties of internet intermediaries at an international level (Nolan, 2013).

Non-binding human rights rhetoric presents an opportunity for advocates to debate, discuss, and particularise the responsibilities of the businesses that make internet communications possible to limit abuse. We take seriously the claim that the UNGPs can help to articulate the balance of competing rights to better delineate the responsibilities of telecommunications companies and digital media platforms. This work has already begun, but so far mostly in relation to freedom of expression and privacy rights. Less attention has been paid to date to the human rights implications

of GBV. In the following sections, we propose some substantive implications which can be drawn from the UNGPs for internet intermediaries to tackle GBV. We use three main themes from the Guiding Principles to structure this analysis: first, that companies should exercise due diligence to identify their impact on human rights; second, that due diligence requires them to take steps to prevent and mitigate negative impact through their policies, procedures, and design choices; and third, that firms have some responsibility to develop effective remedies where human rights have been violated.

Monitoring impact

For internet intermediaries, the first step in fulfilling their responsibility to respect human rights is to identify the impacts that their systems have on the rights of their users and others who may be affected. This implies, firstly, 'that companies have a responsibility to assess the way their practices, services, and products impact on human rights, and to mitigate negative impact' (Jørgensen, 2017, p. 282). The UNGPs require companies to employ due diligence to ensure that their operational policies and procedures not only refrain from creating or facilitating any human rights violations, but also to ensure that they are not complicit in any violations that occur elsewhere in their supply chains, and that human rights-based approaches and considerations are embedded throughout their policies. Impact assessment, performance tracking, and consultation with a variety of stakeholders are required to show that the due diligence standard has been met in this respect (United Nations Human Rights Council, 2011).

Internet intermediaries are starting to develop systems to track their impact on human rights, but these are not yet sufficiently comprehensive or transparent. Some

of this work is being carried out within major internet companies, although the way that these firms understand human rights is often focused on intrusion in expression and privacy by states rather than a detailed understanding of the effects of their own systems and processes (Jørgensen, 2017; Suzor, Van Geelen, Myers West, 2018). This focus can be observed in the way that major telecommunications and internet intermediaries currently report on their performance against human rights standards. Several dozen telecommunications and internet firms now provide regular transparency reports, but the data that is revealed is always selective and partial (Parsons, 2017). Reports from providers are focused on external demands such as government and third-party requests that they remove content or reveal private information rather than the harms they may enable. This is a key area where recognising that GBV is a key human rights issue presents clear guidance and a tool for informal and formal pressure to improve practice: ideally, telecommunications companies and digital media platforms should implement systems that enable them to report on abuse on their networks and the efficacy of systems they have in place to mitigate harm. Telecommunications firms will need to spend considerable time with their diverse stakeholders to understand the range of human rights issues and develop mechanisms to track this impact over time. This is not a straightforward or simple task, but it is a necessary precondition to developing effective responses to GBV and other abuses online.

Mitigating harm through design

Apart from identifying their impacts, the Guiding Principles require platforms to work to mitigate harms that they are involved with. While most digital media platforms

have made public commitments to reducing hate speech and harassment on their platforms in recent years, deeply entrenched structural features of existing social media platforms often exacerbate the effects of online harassment and abuse (Pavan, 2017). For example, Reddit's design and policies around offensive content provided fertile ground for misogynistic activism during the #gamergate controversy (Massanari 2015). The fact that dominant "social media business models and software architecture are oriented towards the maximisation of content sharing and user interaction" can be a key contributing factor to a system that tacitly facilitates widespread abuse and harassment (Salter, 2016, p. 2725). The attention that advertising-driven digital media platforms derive from the spread of highly inflammatory abusive content can create economic disincentives to deal with abuse, although as the recent 'adpocalypse' affecting YouTube shows, advertisers' reluctance to be associated with inappropriate, hateful or abusive content can generate contradictory pressures on platforms to create an environment that is 'safe' for advertisers who do not wish to be associated with hate speech, violence, pornography or political extremism (Gesenhues, 2017). Major platforms also have been unable to articulate clear definitions of acceptable speech; controversial or insensitive humour, for example, is often protected under the Terms of Service by most social media sites, which facilitates pervasive abuse and misogyny disguised as humour (Dragiewicz et al. 2018; Milner 2013; Shifman 2013). For companies that prioritise growth, these pressures are barriers to systematic improvements to responding to GBV and other forms of abuse online. However, Waldman (2018) argues that a combination of law, trust, and community norms can help produce safe social spaces online. Some platforms have built their business models on the demand

for inclusive environments online, designing their products to reduce harm and promote social environments that promote different values.

The Guiding Principles provide insight into the responsibility of platforms to address broader social inequalities that are often reproduced and can exacerbated by their systems. Ultimately, platforms can avoid infringing on the human rights of others by considering human rights throughout the design and operation of their systems. One of the most dangerous engineering practices in this regard is to treat the technology being developed as neutral and, as a consequence, to avoid making considered choices about how potential harmful uses of the tools can be mitigated. Because online GBV reflects the pervasive structural and cultural inequality that exists offline, telecommunications tools that do not actively take inequality into account will inevitably contribute to its amplification. There have been important advances in recent years to more actively understand how design and engineering choices can exacerbate or mitigate underlying inequality, and efforts should be encouraged. At the level of infrastructure, groups like the Human Rights Protocol Considerations Research Group within the Internet Research Task Force have started the difficult work of identifying the human rights implications of technical standards (IETF, 2017a) and convincing those responsible for developing standards that they embed important political choices (DeNardis, 2012; DeNardis and Hackl, 2015; IETF, 2017b). These advances are promising, but a great deal more work remains to be done within technical communities to ensure that internet technologies are designed in a way that empowers women to have full enjoyment of their human rights.

Effective remedies

The UNGP framework includes an emphasis on the development of effective remedies when businesses are involved in human rights violations. Despite repeated commitments to do more to address GBV online, many platforms and telecommunications providers have been slow to improve responses to abuse on their networks. Most digital media platforms rely primarily on relatively simple systems to moderate content -- flagging systems that allow users to identify content for review (Crawford and Gillespie, 2014), and a limited set of blocking and filtering tools that help users manage the material they are exposed to. These systems have so far proven to be deeply inadequate to the task of addressing online abuse at any serious scale (Matias et al., 2015). These systems place responsibility for reporting content on the target of abuse, and they have been ineffective at stemming the tide of widespread abuse (Poland, 2016). Indeed, abuse is effectively normalised on many platforms. Just as is the case elsewhere in social life, tolerating GBV on networks actively reproduces inequality and discrimination against women. When abuse becomes normalised, the socio-technical systems in place to curate content in response to user tastes can entrench what Massanari (2015) calls 'toxic technocultures' by rewarding controversial and anti-social behaviour with higher visibility. Indeed, a pervasive group identity marked by male-dominance may provide an atmosphere which reinforces and then actually elicits harassers' tendencies to disrespect women online (Barak, 2005, p. 82). While it is tempting to blame the internet for the more egregious manifestations of GBV, it is important to remember that these dynamics are not unique to the online context, with patriarchal peer support for abuse of women documented across social contexts (DeKeseredy, Dragiewicz and Schwartz, 2017).

Just as many states have laws stipulating gender equality, most major platforms have clear rules against abusive behaviour. Yet these are inconsistently applied and enforced, often along the lines of existing structural inequality. In order to make meaningful progress on changing user cultures, platforms will need to more systematically respond to abuse and, critically, ensure that these responses are clearly signalled to their users in order to more effectively change the norms of acceptable behaviour.

The task of developing effective responses to online GBV will require the active participation of many different stakeholders. As part of their responsibility to respect human rights, telecommunications providers and digital media platforms will need to develop better complaints systems and mechanisms for due process that are able to effectively mitigate harm perpetrated through their networks. Addressing GBV and other forms of abuse will require platforms to recognise systemic inequality and entrenched power differentials and design their policies and reporting and enforcement processes accordingly. The will be difficult to accomplish without diversification of the tech workforce.

The human rights frame reminds us about the importance of multivalent approaches to developing effective remedies. Ongoing work is needed to address the expression of hatred and intolerance through non-legal measures such as education about the value of human rights, awareness-raising about community conduct standards and the impact of hate speech, efforts to promote counter-speech, and increasing the visibility of civil social dialogue across the social ecology. In order to be effective, these approaches will necessarily require the voluntary participation of grassroots activism and private organisations in national and multinational efforts (La Rue, 2012, Kaye and Šimonović, 2017).

IV Conclusion

Online gender-based violence is increasingly recognised as a major human rights problem. Just like the discriminatory attitudes that engender them, solutions are complex. GBV occurs on a continuum that encompasses a range of behaviours from the routine forms of normalised harassment and abuse to criminal acts. Competing rights and jurisdictional issues will require multifaceted strategies to address GBV. Government intervention alone will not be sufficient to challenge and address systemic inequality, discrimination, and abuse. Ultimately, discriminatory attitudes are produced and reproduced across communities and cultures, and these are what need to change to reduce GBV.

We suggest that the next steps in a multifaceted response to online GBV requires open discussion and debate as well as empirical research to better understand people lived experiences of online GBV. While excellent work is emerging, more research is needed to understand the landscape of GBV and responses to it online. In particular, empirical research is needed in order to produce evidence to guide international efforts to modify and use internet architecture to address GBV. Scholars can investigate the meaning and impact of GBV online as well as the behaviours and technologies involved. Research with internet users and avoiders can help us understand the positive and negative uses of the internet for achieving gender equality and addressing GBV. Ample research documents the gendered digital divide, but it is mostly descriptive and concerned with the percentage of populations accessing the internet, how often, using which platforms, and to a lesser extent, for what. These crude measures fail to capture cultural contributing

factors that are important to understand benefits and barriers to internet use for people of all genders.

In addition (and in parallel) to more research, more experimentation is needed to understand how telecommunications and internet intermediaries can adopt their infrastructure, policies, procedures, and socio-technical affordances to tackle the spectrum of GBV. This work is still in very early stages, and scholars and civil society advocates are leading the way in articulating what societies are entitled to expect from technology firms. Ongoing collaborations between tech firms and civil society have started to bear very promising fruit, but a great deal more work remains to be done. In this regard, while aspirational principles articulated in human rights documents may seem idealistic, we see them as a key part of an ongoing set of efforts to ensure that our globally shared digital infrastructure and the multitude of communications services that build upon it are designed and deployed in a way that respects and promotes human rights. By no means will non-binding principles be sufficient; but as Esquivel and Sweetman argue, "international agreements have been hugely important in directing policy decisions and resource flows to social goods, acting as a rallying cry for those fighting injustice and marginalisation, and influencing the cultural and social norms which we all live by" (2016, p. 2).

We argue that domestic laws, human rights treaties, soft laws, aspirational statements, and multi-stakeholder initiatives all play important roles in addressing GBV online. Nation states retain primary responsibility to develop domestic laws that can effectively protect human rights, and states should take much more urgent action to address what appears to be a rapidly growing problem. Ultimately, however, addressing GBV online requires action by non-state entities, especially corporations based in the most powerful countries. Efforts to enlist these firms in addressing GBV

online face daunting challenges. Conflicting rights and implementation challenges will require collaborative action across business, government, and the broader community, as well as ongoing reform to telecommunications law and policy. This is also an inherently international issue, but international coordination is constrained because different countries have differing levels of commitment to human rights principles, and some have not ratified major treaties. The geographic concentration of these corporations in the US, which has not ratified key human rights treaties such as the Convention on the Elimination of all Forms of Discrimination Against Women (CEDAW), is an undeniable barrier to using human rights tools to address GBV. Nonetheless, in an era of institutionalised disregard for human rights, discussion of gender equality and the elimination of GBV online provides an opportunity to discuss normative values around GBV, including in developed countries. We hope that this discourse can increasingly be used to positively influence the ongoing practices of telecommunications companies and digital media platforms.

References

Amnesty international. 2018. "Toxic Twitter - A toxic place for women." https://www.amnesty.org/en/latest/research/2018/03/online-violence-against-women-chapter-1/.

Association for Progressive Communications. 2017. "Bridging the gender digital divide from a human rights perspective: APC submission to the Office of the High Commissioner for Human Rights." https://www.apc.org/sites/default/files/APCSubmission_OHCHR_BridgingGenderDigitalDivideHumanRightsPerspective_0.pdf.

Athar, R. 2015. "From Impunity to Justice: Improving Corporate Policies to End Technology-Related Violence Against Women." Association for Progressive Communications. https://www.genderit.org/sites/default/upload/flow_corporate_policies_formatted_final.pdf.

Baderin, M., and M. Ssenyonjo, eds. 2016. International Human Rights Law: Six Decades After the UDHR and Beyond. London: Routledge.

Baker, S. 2014. #WhatAreYouDoingAboutVAW campaign: Social media accountability. GenderIT.org. https://www.genderit.org/feminist-talk/whatareyoudoingaboutvaw-campaign-social-media-accountability.

Barak, A. 2005. "Sexual Harassment on the Internet." *Social Science Computer Review* 23 (1): 177-92.

Belli, L., and N. Zingales, eds. 2017. *Platform Regulations: How Platforms are Regulated and how They Regulate us.* Rio de Janeiro: FGV Direito Rio.

boyd, d. 2015. "What World are we Building?" Everett C Parker Lecture, 20 October, Washington, DC.

Bilchitz, D. and S. Deva. 2013. "The human rights obligations of business: a critical framework for the future." In: *Human Rights Obligations of Business: Beyond the corporate responsibility to respect?,* eds. S. Deva and D. Bilchitz. Cambridge: Cambridge University Press. 1–26.

Bittle, S. and L. Snider. 2013. "Examining the Ruggie Report: Can Voluntary Guidelines Tame Global Capitalism?" *Critical Criminology* 21 (2): 177–92.

Chatterjee, R., Doerfler, P., Orgad, H., Havron, S., Palmer, J., Freed, D., Levy, K., Dell, N., McCoy, D. and Ristenpart, T. (2018). The spyware used in intimate partner violence (p. 18). Presented at the IEEE Symposium on Security and Privacy, San Francisco.

Citron, D. K. 2009. "Cyber Civil Rights." *Boston University Law Review* 89: 61–125.

Citron, D. K. 2014. *Hate Crimes in Cyberspace*. Cambridge, MA: Harvard University Press.

Cohen, J. E. 2012. *Configuring the Networked Self: Law, Code, and the Play of Everyday Practice*. New Haven, CT: Yale University Press.

Council of Europe. 2017. Recommendation CM/Rec(2017x)xx of the Committee of Ministers to member states on the roles and responsibilities of internet intermediaries. Third revised draft. January 12. https://rm.coe.int/recommendation-cm-rec-2017x-xx-of-the-committee-of-ministers-to-member/1680731980.

Crawford, K., and T. Gillespie. 2014. "What is a Flag for? Social Media Reporting Tools and the Vocabulary of Complaint." *New Media & Society* 18 (3): 1–9.

Croeser, S. 2016. "Thinking Beyond Free Speech' in Responding to Online Harassment', *Ada: a journal of new media, gender, and culture*, 10.

DeKeseredy, W.S., M. Dragiewicz, and M.D. Schwartz. 2017. *Abusive endings: Separation and divorce violence against women*. Oakland, CA: University of California Press.

DeNardis, L. 2012. "Hidden Levers of Internet Control." *Information, Communication and Society* 15 (3): 720-38.

DeNardis, L., and A.M. Hackl. 2015. "Internet Governance by Social Media Platforms." *Telecommunications Policy* 39 (9): 761-70.

Dibbell, J. 1994. "A rape in cyberspace or how an evil clown, a Haitian trickster spirit, two wizards, and a cast of dozens turned a database into a society." *Annual Survey of American Law* 3: 471-89.

Dragiewicz, M., J. Burgess, A. Matamoros-Fernández, M. Salter, N.P. Suzor, D. Woodlock and B. Harris. 2018. "Technology facilitated coercive control: domestic violence and the competing roles of digital media platforms." *Feminist Media Studies* 1-17.

Esquivel, V., and C. Sweetman. 2016. "Gender and the Sustainable Development Goals." *Gender & Development* 24 (1): 1–8.

Franks, M.A. 2011. "Unwilling Avatars. Idealism and Discrimination in Cyberspace." *Columbia Journal of Gender and Law* 20: 224-61.

Fyfe, S. (2017). Tracking hate speech acts as incitement to genocide in international criminal law. *Leiden Journal of International Law* 30 (2): 523–48.

Gesenhues, A. 2017. "YouTube: We've Manually Reviewed 1M+ Videos to Improve Brand Safety Processes." *Marketing Land.*

https://marketingland.com/youtube-manually-reviewed-1m-videos-improve-brand-safety-efforts-around-extremist-content-226579

Gillespie, T. 2017. "Governance of and by platforms." In *SAGE Handbook of Social Media,* eds. J. Burgess, T. Poell, and A. Marwick. Thousand Oaks: SAGE Publications.

Gray, D.C., D.K. Citron, and L.C. Rinehart. 2013. Fighting cyber-crime after United States v. Jones. *The Journal of Criminal Law & Criminology*, *103*(3), 745–802.

Hasselbacher, L. 2010. "State Obligations Regarding Domestic Violence: The European Court of Human Rights, Due Diligence, and International Legal Minimums of Protection." *Northwestern Journal of International Human Rights* 8 (2): 190-215.

Haque, U. 2016. "The Reason Twitter's Losing Active Users." Harvard Business Review. *https://hbr.org/2016/02/the-reason-twitters-losing-active-users*

Henry, N., and A. Powell. 2016. "Sexual Violence in the Digital Age: The Scope and Limits of Criminal Law." *Social & Legal Studies* 25(4): 397–418.

Hussain, A., 1997. Question of the Human Rights of All Persons Subjected to Any Form of Detention or Imprisonment Promotion and Protection of the Right to Freedom of Opinion and Expression (No. E/CN.4/1997/31). UN Commission on Human Rights, Geneva.

Hussain, A. 1998. *Question of the Human Rights of All Persons Subjected to Any Form of Detention or Imprisonment Promotion and Protection of the Right to Freedom of Opinion and Expression: Report of the Special Rapporteur*. Geneva: UN Economic and Social Council.

IETF. 2017a. "Guidelines for Human Rights Protocol Considerations." January 12. https://datatracker.ietf.org/doc/draft-tenoever-hrpc-guidelines/?include_text=1.

IETF. 2017b. "On the Politics of Standards." https://datatracker.ietf.org/doc/draft-tenoever-hrpc-political/

IGF (Internet Governance Forum). 2015. "Online Abuse and Gender-Based Violence Against Women." https://www.intgovforum.org/multilingual/es/content/online-abuse-and-gender-based-violence-against-women

Jane, E.A. 2017. Misogyny Online: A Short (and Brutish) History. London: SAGE Publications.

Jørgensen, R.F. and A.M. Pedersen. 2017. "Online Service Providers as Human Rights Arbiters." In *The Responsibilities of Online Service Providers*. (eds) M. Taddeo and L. Floridi. Springer: London. 179–99.

Jørgensen, R.F. 2017. "What Platforms Mean When They Talk About Human Rights." *Policy & Internet* 9 (3): 280-96.

Kaye, D. 2016. *Report of the the Special Rapporteur to the Human Rights Council on Freedom of expression, states and the private sector in the digital age*, A/HRC/32/38. https://documents-dds-ny.un.org/doc/UNDOC/GEN/G16/095/12/PDF/G1609512.pdf?OpenElement

Kaye, D. and D. Šimonović. 2017. "UN experts urge States and companies to address online gender-based abuse but warn against censorship." 8 March. 2017http://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=21317&LangID=E

Kelly, L. 1988. *Surviving Sexual Violence.* Oxford: Polity Press.

Laidlaw, E.B. 2017. "Myth or Promise? The Corporate Social Responsibilities of Online Service Providers for Human Rights." In *The Responsibilities of Online Service Providers*, eds. M. Taddeo and L. Floridi. Springer: London. 135-55.

La Rue, F. 2012. *Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*. A/67/357 http://ap.ohchr.org/documents/dpage_e.aspx?si=A/67/357.

Lenhart, A., M. Ybarra, K. Zickuhr and M. Price-Feeney. 2016. *Online Harassment, Digital Abuse, and Cyberstalking in America*. New York: Data & Society Research Institute, Centre for Innovative Public Health Research.

MacKinnon, R., E. Hickock, A. Bar and H. Lim. 2014. *Fostering freedom online: the role of internet intermediaries*. Paris: UNESCO.

Mares, R., ed. 2012. *The UN Guiding Principles on Business and Human Rights: Foundations and Implementation*. Leiden: Brill.

Massanari, A. 2015. "#Gamergate and the Fappening: How Reddit's Algorithm, Governance, and Culture Support Toxic Technocultures." *New Media & Society* 19 (3): 329-46.

Matias, J.N., A. Johnson, W.E. Boesel, B. Keegan, J. Friedman, and C. DeTar. 2015. "Reporting, Reviewing, and Responding to Harassment on Twitter." Women, Action, and the Media. http://womenactionmedia.org/twitter-report.

McConnell, L. 2017. "Assessing the Feasibility of a Business and Human Rights Treaty." *International and Comparative Law Quarterly* 66 (1): 143–80.

McQuigg, R.J.A. 2011. *International Human Rights Law and Domestic Violence: The Effectiveness of International Human Rights Law.* London: Routledge.

Milner, R.M. 2013. "FCJ-156 Hacking the Social: Internet Memes, Identity Antagonism, and the Logic of Lulz." *The Fibreculture Journal*. 156 (22): 62-92.

Napoli, P. 2001. *Foundations of communications policy: Principles and process in the regulation of electronic media*. Cresskill, NJ: Hampton Press.

Nolan, J. 2013. "The Corporate Responsibility to Respect Rights: Soft Law or Not Law?" In: *Human Rights Obligations of Business: Beyond the Corporate Responsibility to Respect?,* eds. S. Deca and D. Bilchitz. Cambridge: Cambridge University Press, 1-27.

OECD (Organisation for Economic Cooperation and Development). 2010. *The Economic and Social Role of Internet Intermediaries*. Geneva: Organisation for Economic Cooperation and Development. https://www.oecd.org/internet/ieconomy/44949023.pdf.

OHCHR (Office of the United Nations High Commissioner for Human Rights). 2017. "Call for Submission on Online Violence Against Women." 8 January. http://www.ohchr.org/EN/Issues/Women/SRWomen/Pages/OnlineViolence.aspx.

Parsons, C. 2017. "The (In)effectiveness of Voluntarily Produced Transparency Reports." *Business and Society* 1-29.

Pavan, E. 2017. "Internet Intermediaries and Online Gender-Based Violence." In *Gender, Technology and Violence*, eds. M. Segrave and L. Vitis. London: Routledge, 62-78.

Poland, B. 2016. *Haters: Harassment, Abuse, and Violence Online*. Lincoln: University of Nebraska Press.

Potter, H. 2015. *Intersectionality and criminology: Disrupting and revolutionizing studies of crime*. London: Routledge.

Ranking Digital Rights. 2017. '2017 Corporate Accountability Index.' *Ranking Digital Rights*. 8 Jan. https://rankingdigitalrights.org/index2017/.

Richie, B. E. 2012. *Arrested justice: Black women, violence, and America's prison nation*. New York: NYU Press.

Salter, M. 2016. "Privates in the Online Public: Sex(ting) and Reputation on Social Media." *New Media & Society* 18 (11): 2723-39.

Shifman, L. 2013. *Memes in Digital Culture*. Cambridge, MA: The MIT Press.

Šimonović, D., 2016. *Report of the Special Rapporteur on Violence Against Women, Its Causes and Consequences* (No. A/HRC/32/42). UN Human Rights Council, Geneva.

Stanko, E.A. 1990. *Everyday Violence: How Women and Men Experience Sexual and Physical Danger.* London: Pandora.

Streeter, T. 2011. *The Net Effect: Romanticism, Capitalism, and the Internet.* New York: New York University Press.

Suzor, N., T. Van Geelen, and S. Myers West. 2018. "Evaluating the legitimacy of platform governance: A review of research and a shared research agenda." *International Communication Gazette* 80 (4): 385-400.

Suzor, N., B. Seignior and J. Singleton. 2017. "Non-consensual porn and the responsibilities of online intermediaries." *Melbourne University Law Review* 40 (3): 1057-97.

Turner, F. 2006. *From Counterculture to Cyberculture: Stewart Brand, the Whole Earth Network, and the Rise of Digital Utopianism*. Chicago: University of Chicago Press.

Turton-Turner, P. 2013. "Villainous avatars: The visual semiotics of misogyny and free speech in cyberspace." *Forum on Public Policy* 1–18.

UN Broadband Commission for Digital Development Working Group on Broadband and Gender. 2015. Cyber Violence Against Women and Girls. UNDP.

http://www.broadbandcommission.org/Documents/reports/bb-wg-gender-discussionpaper2015-executive-summary.pdf

United Nations. 2015. Transforming Our World: The 2030 Agenda for Sustainable Development. Geneva: United Nations. https://sustainabledevelopment.un.org/post2015/transformingourworld.

United Nations Human Rights Council. 2011. Guiding Principles on Business and Human Rights. http://www.ohchr.org/documents/publications/GuidingprinciplesBusinesshr_en.pdf

Waldman, A. E. 2018. "Safe social spaces." *NYLS Legal Studies Research Paper No. 3119735*.

Wired Staff. 2016. "Dear Internet: It's Time to Fix This Mess You Made." WIRED. https://www.wired.com/2016/08/open-letter-to-the-Internet/ January 8.

Venturini, J., L. Louzada, M. Maciel, N. Zingales, K. Stylianou, L. Belli. 2016. *Terms of Service and Human Rights: An Analysis of Online Platform Contracts.* Rio de Janeiro: Editora Revan.