Human Rights Responsibilities of Platforms for Addressing Gender-based Violence Online

I. The internet has an abuse problem

The internet has a widely recognised abuse problem. Indeed, in late 2016, the editorial staff at the long-established technology publication *WIRED Magazine* wrote an open letter to the internet that explicitly acknowledged how abuse and harassment have 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 to 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 the engineers who design and administer internet protocols, platforms, and services have an obligation to protect users from abuse by private individuals or collectives of individuals.

1

In this article, we try to understand the role of internet businesses in combating online abuse. We focus on internet intermediaries -- telecommunications providers, search engines, content hosts, social media platforms -- that play a major role in mediating communications between internet users (OECD, 2010). The decisions these intermediaries make in governing their networks have direct impact on the 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 social life, intermediaries have moral, 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). The exact scope of those responsibilities, however, is still heavily contested.

Telecommunications corporations and digital media platforms have been slow to respond to online abuse and harassment, persisting in representing themselves as neutral intermediaries. There is little consensus about what actions they are expected to 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 abusively, nor are they under any substantial obligations to design their networks in ways that limit abuse. In the absence of clear legal obligations, these intermediaries have little guidance about how they should respond.

In this article, we argue that telecommunications providers and internet intermediaries have human rights responsibilities to design their networks and services to combat abuse. In Part II, we explain how gender-based violence (GBV) encompasses a continuum of harm that can only be understood in the context of underlying structural inequalities (Kelly, 1988). Online GBV ranges from the use of technology as a tool of coercive control within intimate relationships to the widespread misogyny that has become normalised on the internet today. Addressing online GBV, then, requires the active and careful participation of state and non-state actors involved in internet governance.

In Part III, we examine how human rights norms may help to better inform ongoing debates about the responsibilities of digital media platforms to address GBV on their networks. While the primary obligations to protect human rights rest with nation states, the obligations of states must be understood in the context of a diverse, decentralised internet whose governance is dominated by private entities -- and particularly by major national and transnational telecommunications firms and digital media platforms. The decentralised nature of internet architecture and governance requires a more integrated response. We argue that an environment that enables people to exercise their human rights online can only be created with the active participation of the engineers and operators of internet intermediaries. This will inevitably require difficult and ongoing assessments of the role of technology in social interactions and the appropriate mechanisms to promote human rights.

The UN Guiding Principles on Business and Human Rights articulate clear expectations that platforms have a responsibility to design and operate their networks in a way that avoids infringing on the rights of those they affect, as well as to find new, effective remedies for online abuse. There is as yet little consensus on the substance of these responsibilities, but we suggest that the language of responsibility can help progress debates about what actions telecommunications companies and digital media platforms should be expected to take to ensure that the networks they create do not reproduce gendered inequality. Importantly, while the language of responsibility can help to identify expectations of internet intermediaries, the Guiding Principles are not binding, and may not be sufficient to drive the cultural and technical changes that are needed to address gender inequality. The ongoing development of a treaty on business and human rights presents one promising option to create binding obligations. Meanwhile, however, 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 initial recommendations for both state and non-state actors to better address gender-based violence online. For states to fulfil their obligation to protect human rights, we suggest first that online abuse needs to be taken more seriously than it has to date by law enforcement agencies. We also note the need for immediate 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. As to the substance of what businesses can do, we argue that we are only at the beginnings of necessary global conversations about how to address gender inequality online. There are, however, a set of clear initial steps for businesses, including particularly (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 violence 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 a potential revitalisation of substantive work to advance efforts to address GBV online as a human rights issue. Importantly, this agenda specifically seeks to examine the obligations of states in prosecuting online violence against women as well as the role of internet intermediaries in identifying, reporting, and rectifying incidents of harassment or violence (OHCHR, 2017). This is a critical point: in this Part, we argue that GBV can only be addressed systematically, with the involvement of both state and non-state actors, and by targeting pervasive, commonly accepted abuse as well as those acts that are clearly recognised as criminal.

In order to understand the extent and implications of online GBV, abuse and harassment must be understood in the context of a broader landscape of inequality and harm (DeKeseredy, Dragiewicz and Schwartz, 2017). Kelly (1988) proposed that multiple forms of abuse and violence against women comprise a continuum, where 'typical' and 'aberrant' male behaviours shade into one another. In other words, the continuum of violence ranges from behaviours which are normalised (including widespread misogyny and commonplace sexism) to those widely recognised as criminal or deviant. Importantly, the 'continuum' in this context specifically refers to a range of prevalence, not seriousness. In this framing, different acts, from harassment to physical violence, should not be understood as more or less serious. Instead, Kelly highlights the common threads linking abusive acts across an array of settings throughout women's lives. Her concern is with the ways these acts serve to limit the lives and agency of women and their 'space for action'. Abuses that women are subjected to at home, work, and leisure can create 'climates of unsafety' (Stanko, 1990) which threaten the rights and freedoms of women. Similarly, widespread, dispersed online harassment and abuse directly threatens the autonomy, dignity, and privacy of targets of abuse and constrains their ability to participate in contemporary social life (Citron, 2014). Along this continuum, different types of gender-based violence are all profoundly shaped by gender

inequality, and are all underpinned by common social attitudes that minimise and excuse violence against women.

Adopting the continuum, we argue that online and offline contexts are essential to understanding GBV and securing women's enjoyment of human rights. Key stakeholders have recognised the importance of online abuse for women. For example, the Association for Progressive Communications (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).

Gender-based violence is a both a cause and effect of inequality and a barrier to social and economic development. The threat of violence and sexual abuse can, for example, prevent civic engagement and pursuit of education, limiting social participation and employment opportunities. Structural inequality makes women vulnerable to violence and hinders their abilities to exit abusive relationships. This conceptualisation of GBV on a continuum is compatible with emerging priorities within international institutional discourse; 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). Targets 5.1 ('End all forms of discrimination against all women and girls everywhere') and 5.2 ('Eliminate all forms of violence against all women and girls in the public and private sphere') explicitly link GBV and inequality.

The clearest human rights obligations to address GBV online arise from the acts commonly recognised as aberrant -- acts that are either directly criminalised or closely linked to criminal behaviour. One of the key areas of concern is when technology is used to enact harm in the context

of domestic violence, increasingly referred to as coercive control (Stark, 2007; Woodlock, 2016). Intimate partners use physical, psychological, sexual and/or financial abuse and, increasingly, technology to exercise and amplify abuse, in efforts to coerce, control and threaten women (Authors, 2018; Woodlock, 2016). This phenomenon, 'technology-facilitated coercive control', includes abusive intimate partners using a diverse range of technologies to harass, threaten, monitor, stalk, impersonate and dox victim/survivors (Authors, 2018). We emphasise that such behaviours are not isolated from but embedded within a woman's experience of coercive and controlling behaviours and threats of violence perpetrated 'offline'. These forms of GBV are inextricably contextualised in relationship dynamics, culture and structural inequality (see Stark, 2007; Authors, 2018).

The use of telecommunications technologies as tools of coercion and control by intimate partners as part of domestic violence clearly raises substantive human rights concerns. The United Nations Declaration on the Elimination of Violence against Women defines violence against women as physical, sexual and psychological violence occurring in the family, and 'perpetrated or condoned by the State, wherever it occurs' (OHCHR, 1993, Article 2). Domestic violence often constitutes a violation of the right to be free from degrading or inhuman treatment, the right to privacy or private life (on the basis of the interference with dignity and identity), the right to bodily integrity, and, in the most extreme cases, the right to life (McQuigg, 2011, xiii; Hasselbacher, 2010, 191; Bettinger-Lopez, 2008; Ewing, 1994; Charlesworth, 1994; *Opuz v Turkey*; *Lenahan v U.S.*). Because domestic violence disproportionately impacts women and minorities, it is also often considered to be a violation of the right to non-discrimination (Hasselbacher, 2010; Hanna, 2011; Rudolf and Eriksson, 2007; Ewing, 1994).

States have clear human rights obligations to protect women from gender-based violence perpetrated through telecommunications technologies. In part, this means greater enforcement for criminal behaviour. International human rights law in domestic violence cases clearly establish the principle that states have a direct responsibility to take effective action to exercise due diligence to

prevent or respond to violations of human rights by private parties (McQuigg, 2011). One of the major persistent problems faced by victims of online abuse is that threats made against them are not taken seriously by law enforcement authorities -- state agents still commonly view technology-facilitated abuse and stalking as less serious than or separate from 'traditional' violence (George and Harris, 2014, Harris, 2016). Law enforcement are more likely to respond to prolonged and focused harassment than individual incidents, and some agencies are developing specialist capabilities to respond to online abuse, but state responses remain inconsistent at best.

Importantly, the obligation to protect requires not only that states effectively enforce criminal law, but also that they create an environment where people are able to exercise their rights. Understanding online abuse and harassment on a continuum of gender-based violence highlights the importance of paying increased attention to normalised, everyday acts of abuse that work to limit the participation of women online. These seemingly 'less serious' acts cannot be divorced from acts that are widely recognised as causing harm (Vera-Gray, 2014). Practices of online harassment such as offensive name calling, trolling, bullying, threats, impersonation, the unauthorised production and/or sharing of intimate images are both rampant and the subject of significant popular, scholarly and policy attention (see Franks, 2012; Lenhart, Zickuhr and Price-Feeney, 2016; Weinstein and Selman, 2014). While some of these practices are native to the internet and digital media, others pre-exist but are transformed by these technologies and platforms. Online harassers use a variety of communications technologies and platforms to harass women, but regardless of the technology used, online harassment generally assumes sexualised and/or misogynistic elements. It can include patterns or sustained campaigns and particular targets (such as in the 2014 #gamergate controversy, which saw harassment of women in the video game industry, see Massanari, 2015; Salter, 2017b) and also isolated incidents by unknown or known perpetrators (and indeed some use the anonymity and the clandestine environments of various platforms in effort to hide their identity). It also includes less organised and more dispersed everyday abuse that has become accepted online -- what Jane (2012:

532) terms 'e-bile': 'extravagant invective, the sexualized threats of violence, and the recreational nastiness that have come to constitute a dominant tenor of Internet discourse'.

We suggest that one of the major reasons that online GBV against women is tolerated or overlooked is because the nature and impact of digital harms are not widely understood. Online harassment, for example, is often excused as harmless 'trolling' (Turton-Turner, 2013). The use of freedom of speech as a key framing concept in discourse about the governance of online environments operates to delegitimise claims of harm as a result of online harassment and abuse (Croeser, 2016). The roots of this dismissal of online harassment as mere speech can be traced to the early days of the commercial internet, where 'cyberspace' was imagined as a different place, and its potential to contribute to libertarian ideals was emphasised (Cohen, 2012). But this utopian vision obscures the internet's potential to further embed and amplify discrimination (Franks, 2011).

The conflict between protecting freedom of speech online and preventing abuse is a false dichotomy, and one which has been recognised for a long time in the international human rights discourse. The UN Special Rapporteurs on the Promotion and Protection of the Right to Freedom of Opinion and Expression have a twenty-year history of noting the human rights implications of discrimination, freedom of expression and privacy for women, including in relation to GBV. Indeed, the Commission on Human Rights resolution 1997/27 included a mandate to:

> pay particular attention to the situation of women and the relationship between the effective promotion and protection of the right to freedom of opinion and expression and incidents of discrimination based on sex, creating obstacles for women with regard to their right to seek, receive and impart information, and to consider how those obstacles impeded the ability of women to make informed choices in areas of particular importance to them, as well as in areas related to the general decision-making processes in the societies in which they live. (Hussain, 1998, §E. 49)

Hussain urged 'Governments to ensure that effective measures are taken to eliminate the atmosphere of fear that often prevents many women from communicating freely on their own behalf or on behalf of other women who have been victims of violence either in domestic or community settings or as a result of internal or trans-border conflict' (Hussain, 1997, §62). He expressed a desire to work with the Special Rapporteur on Violence against Women to 'identify the obstacles to freedom of opinion and expression for women but to establish a framework through which violations of women's right to freedom of expression can be systematically documented and addressed' (Hussain, 1997, §63). Hussain further noted that states' protection of women's rights to expression is a reflection of their national 'standard of fair play, justice and honesty with regard to women and the status accorded to them in society' (Hussain, 1998, §E. 50). His later reports reiterated these concerns and expressed continuing regret that he had been unable to coordinate work with the Special Rapporteur on Violence Against Women (Hussain, 1999). Hussain urged 'Governments to take all necessary steps to remove formal and cultural obstacles to the exercise by women of their right to freedom of expression, including to receive information, and ultimately to give effect to all their rights' to participation in public life (Hussain, 2000, §325). Separately, he expressed concern about new technologies and their implications for human rights and development, calling on states to 'ensure that there is adequate and unfettered "social and political space" in which the new technologies can be developed in a self-regulating environment and where the exercise and enjoyment of the rights to expression, opinion, information, association and assembly can flourish' (Hussain, 2000, §69). These reports provide an early foundation for addressing GBV online as a condition for global development and exercise of human rights. More recently, efforts to collaborate with the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression have finally come to fruition. Šimonović and Kaye issued a joint statement on International Women's Day on 8 March 2017 calling for systematic

investigation of online GBV and remedies that maintain robust protection of rights to privacy and expression (OHCHR, 2017).

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. This aligns with the more recent 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".

The United Nations human rights infrastructure provides a potentially powerful but as yet under-utilised tool for identifying strategies to combat online abuse. Human rights frameworks provide a foundation for basic rights and protections that is an important counterpoint to the false equivalences around free-expression promoted by neo-liberal and libertarian logics. Human rights prohibitions against discrimination, hate, and violence make clear that rights to freedom of expression and conduct carry associated responsibilities not to impinge on the rights of others.

Online abuse is not an issue that can be addressed by states alone through the criminalisation and enforcement of the most aberrant 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. Achieving real change will require the active participation of a wide range of other social institutions to create an online environment where people are free to exercise their rights. In the next Part, we argue that one of the key advantages of using the human rights frameworks in this context is that it provides a mechanism to identify the role of different actors in responding to and preventing abuse. In particular, it highlights that while the state has the primary duty to protect human rights, private actors have responsibilities to mitigate violations of human rights in which they

are involved. This, we suggest, provides a way to understand the responsibilities of the businesses that make internet communications possible may have to design their systems to limit abuse.

III The responsibilities of telecommunications companies and digital media platforms

Human rights law has historically focused primarily on the obligations of state actors. The traditional public/private divide enshrined in the early international human rights regime (McQuigg, 2011; Baderin and Ssenyonjo, 2016) has meant that it has not been easy to understand in any detail what role private actors should play in promoting human rights (Hasselbacher, 2010). The recent United Nations Guiding Principles on Business and Human Rights (UNGPs) outlined 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). The UNGPs demand that states clearly articulate their expectations that all businesses under their jurisdiction respect human rights. The operational principles require that states:

(a) Enforce laws that are aimed at, or have the effect of, requiring business enterprises to respect human rights, and periodically to assess the adequacy of such laws and address any gaps;

(b) Ensure that other laws and policies governing the creation and ongoing operation of business enterprises, such as corporate law, do not constrain but enable business respect for human rights;

(c) Provide effective guidance to business enterprises on how to respect human rights throughout their operations;

(d) Encourage, and where appropriate require, business enterprises to communicate how they address their human rights impacts. (United Nations Human Rights Council, 2011, 3)

These principles begin to map out the division of responsibility between states and businesses. They encourage states to educate businesses about human rights and encourage or require their participation in upholding them. They also make clear that neither states nor businesses alone can fully realise human rights goals.

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. In recent years, a number of different reports from international human rights institutions have started to explicitly call on internet intermediaries to do more to protect the rights of their users (see, e.g. Council of Europe, 2017; MacKinnon et al., 2014; Kaye, 2016). 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).

These initiatives point towards a fruitful way to understand the responsibilities of telecommunications companies and digital media platforms to promote human rights, but so far most of the work has been conducted in relation to freedom of expression and privacy rights. Less

attention has been paid to date to the human rights implications of GBV, and there is a great deal of work to be done to understand how different types of internet intermediaries should respond.

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 '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 only recently starting to develop systems to track their impact on human rights, and these are generally not yet sufficiently comprehensive or transparent. Some of this work is undoubtedly being carried out within major internet companies, although the way that these firms understand human rights is generally skewed towards interference by states, rather than a detailed understanding of the effects of their own systems and processes (Jørgensen, 2017). This can particularly be seen in the way that major telecommunications and internet intermediaries report on their performance against human rights standards. Several dozen telecommunications and internet firms now provide regular transparency reports, although the data that is revealed is always selective and partial (Parsons, 2017). Among all of these reports, however, providers are only reporting on external demands -- particularly government and third-party requests that they remove

content or reveal private information. None of the providers release clear data about the impact of their own content moderation systems, and the focus of these reports is overwhelmingly on state interference with speech and privacy rights (Suzor, Van Geelen, Myers West, 2017). This is a key area where recognising that GBV is a key human rights issue presents clear guidance: ideally, telecommunications companies and digital media platforms should immediately begin to implement systems that enable them to report on abuse on their networks and the efficacy of systems they have in place to mitigate harm. If this is to be done, telecommunications firms will need to spend considerable time with their stakeholders to understand the range of their impact on rights, 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 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. This is a major unsolved challenge. While most digital media platforms have made public commitments to reducing hate speech and harassment on their platforms in recent years, actual progress has been limited. There are deeply entrenched structural features of existing social media platforms that operate to exacerbate the effects of online harassment and abuse. The fact that "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: 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 pressure on platforms to create an environment that is 'safe' for advertisers (Gesenhues, 2017). Authors (2018) note that

contentious, controversial or sensitive humour is often protected under the Terms of Service by most social media sites, which facilitates pervasive abuse and misogyny disguised as humour (see also Milner, 2013; Shifman, 2013). For companies whose ongoing survival depends upon growth, this can create a path dependency where large sections of the user community come to expect that abuse will not be dealt with in any systematic way.

The Guiding Principles provide some insight into the responsibility of platforms to address underlying social inequalities that are enabled and exacerbated by their systems, but there is no easy solution to understanding how this can be done. 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 abuse reflects deep social inequalities, telecommunications tools that do not actively take inequality into account will almost inevitably contribute to the amplification of inequality. There have been important advances in recent years to more actively understand how design and engineering choices can exacerbate underlying inequality, and these moves 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 those standards embed important political choices (DeNardis, 2012; DeNardis and Hackl, 2015; IETF, 2017b). Major platforms, meanwhile, have established internal working groups and partnerships with key NGO and other stakeholders to improve their safety features -- although the effectiveness of these arrangements is not easy to evaluate from external observation. A great deal more work remains to be done, however, to ensure that contemporary communications tools work to empower

women rather than continue to exacerbate well-established patterns of inequality and abuse, or to prioritise the safety of advertisers' brands over the safety of the user community.

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, platforms and telecommunications providers continue to struggle to respond effectively to abuse. Modern 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 manifestly 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 effective at stemming the tide of widespread abuse. Instead, abuse has become normalised on many platforms. Ultimately, 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: 82).

While most major platforms have clear rules against abusive behaviour, these are inconsistently applied and enforced. 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. The human rights frame also suggests the importance of a variety of approaches to develop effective remedies. In particular, much more work needs to be done 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. In order to be effective, these approaches will necessarily require the voluntary participation of private organisations in national and multinational efforts (La Rue, 2012, Kaye and Šimonović, 2017).

Importantly, however, there are limits to a non-binding responsibility approach that need to be dealt with in order to achieve real social change. Businesses face real, practical incentives that often conflict with their ability or willingness to respect human rights (Bittle and Snider, 2013; McConnell, 2017). 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, we might expect states to immediately do more to explore when these responsibilities should be turned into legal obligations under domestic law. In the longer term, the ongoing efforts to develop a binding treaty

on business and human rights provide a promising avenue to better articulate the legal duties of internet intermediaries at an international level.

IV Conclusion

Scholarship on efforts to grapple with the human rights implications of the internet points to foundational principles to open a more comprehensive conversation about the role of platforms in responding to GBV online. Human rights reports and instruments provide one tool for promoting necessary conversations about rights and responsibilities shared across individuals, corporations, and states. 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 criminal remedies may have unintended consequences (LaRue, 2012). Accordingly, the Special Rapporteur advocates tightly tailored laws with high thresholds and clear definitions, state application of civil rather than criminal penalties in most cases, clarification of community values, and more rather than less speech. Research on privacy and free expression is equally difficult and important as efforts to address GBV online. Efforts to enlist platforms in addressing GBV online face daunting challenges. Conflicting rights and implementation challenges will require collaborative action across business, government, and the broader community. 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.

While human rights frameworks provide a promising set of potential tools, little is still known about the landscape of GBV online. Empirical research is needed in order to produce evidence to guide international efforts to 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. However, 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.

Aspirational principles articulated in human rights documents may seem idealistic, 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: 2). Accordingly, their importance should not be underestimated. We argue that human rights treaties, soft laws, and multi-stakeholder initiatives all play important roles in addressing GBV online, although we note that nation states retain primary responsibility to develop domestic laws that can effectively protect human rights.

Ultimately, addressing GBV online requires action by non-state entities, especially corporations based in the most powerful countries. The 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 the era of Trumpism and 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 the most 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

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

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

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.

Bettinger-Lopez, C. 2008. "Human Rights at Home: Domestic Violence as a Human Rights Violation." *Columbia Human Rights Law Review* 40: 19-77.

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

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

Charlesworth, H. 1994. "What are "Women's International Human Rights"?" In *Human Rights of Women: National and International Perspectives,* ed. R.J. Cook. Philadelphia: University of Pennsylvania Press, 58-84.

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.

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

Ewing, A.P. 1994. "Establishing State Responsibility for Private Acts of Violence Against Women Under the American Convention on Human Rights." *Columbia Human Rights Law Review* 26: 751-800.

Franks, M.A. 2012. "Sexual Harassment 2.0." *Maryland Law Review* 71 (3): 655-704.

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

George. A., and B. Harris. 2014. *Landscapes of Violence: Women Surviving Family Violence in Regional and Rural Victoria*. Victoria, Deakin University.

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, T., and A. Marwick. Thousand Oaks: SAGE Publications.

Hanna, C. 2011. Health, Human Rights, and Violence Against Women and Girls: Broadly Redefining Affirmative State Duties after *Opuz v. Turkey*. *Hastings International and Comparative Law Review* 34: 127–9.

Harris, B. 2016. "Violent Landscapes: A Spatial Study of Family Violence." In *Locating Crime in Context and Place: Perspectives on Regional, Rural and Remote Australia,* eds. A. Harness, B. Harris, and D. Baker. Sydney: The Federation Press, 70-84.

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. 1999. *Report of the Special Rapporteur on the Right to Seek and Receive Information, National Security Laws, Criminal Libel, New Information Technologies, and Women and Freedom of Expression*. Geneva: UN Economic and Social Council.

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.

Hussain, A. 2000. *Civil and Political Rights, Including the Question of Freedom of Expression: Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*. Geneva: UN Commision on Human Rights.

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. 2012. ""Your a Ugly, Whorish, Slut": Understanding E-bile." *Feminist Media Studies* 14 (4): 531-46

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

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

LaRue, 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.

Lenahan v. United States. 2011. Case 12.626, Inter-American Commission on Human Rights, Report No. 80/11, 160-62, 172.

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): 329-346.

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

  Manila Principles on Intermediary Liability: Best Practices Guidelines for Limiting Intermediary Liability for Content to Promote Freedom of Expression and Innovation (2015). Manila, Philippines. Available at: https://www.manilaprinciples.org/ (Accessed: 15 February 2017).

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.

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). 1993. Declaration
    on the Elimination of Violence against Women.

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

Opuz v. Turkey. 2009. "Opuz and Interights (intervening) v Turkey, Merits and just satisfaction,
    App No 33401/02, [2009] ECHR 870, (2010) 50 EHRR 28, 27 BHRC 159, IHRL 3618
    (ECHR 2009), 9th June 2009, European Court of Human Rights [ECHR]"
    http://opil.ouplaw.com/view/10.1093/law:ihrl/3618echr09.case.1/law-ihrl-3618echr09.

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

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

Rudolf, B., and A. Eriksson. "Women's Rights Under International Human Rights Treaties: Issues
    of Rape, Domestic Slavery, Abortion, and Domestic Violence." *International Journal of
    Constitutional Law* 5 (3): 507-25.

Salter, M. 2017. "From Geek Masculinity to Gamergate: The Technological Rationality of Online
    Abuse." *Crime, Media, Culture* 1-18.

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 a Digital World: Reconciling With a Conceptual Troublemaker."
    Journal of Computer-Mediated Communication 18 (3): 362-77.

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.

Stark, E., 2007. *Coercive Control: How Men Entrap Women in Personal Life*. New York: Oxford University Press.

Suzor, N., Van Geelen, T., Myers West, S., 2017. Evaluating the legitimacy of platform governance: A review of research and a shared research agenda. International Communication Gazette.

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

United Nations Special Rapporteur on Freedom of Opinion and Expression, the Organization for Security and Co-operation in Europe (OSCE) Representative on Freedom of the Media, the Organization of American States (OAS) Special Rapporteur on Freedom of Expression and the African Commission on Human and Peoples' Rights (ACHPR) Special Rapporteur on Freedom of Expression and Access to Information. 2011. "Joint Declaration on Freedom of Expression and the Internet." http://www.osce.org/fom/78309?download=true

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.

Woodlock, D. 2016. "The Abuse of Technology in Domestic Violence and Stalking." *Violence Against Women* 23 (5): 584-602.

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.

Vera-Gray, F. 2014. *The Great Problems are in the Street: A Phenomenology of Men's Stranger Intrusions on Women in Public Space.* Thesis, DPhil, London Metropolitan University.

Weinstein, E.C., and R.L. Selman. 2016. "Digital Stress: Adolescents' Personal Accounts." *New Media and Society* 18 (3): 391-409.