REGULATING HATE CRIME IN THE DIGITAL AGE

Chara Bakalis

INTRODUCTION

The questions surrounding the regulation of online hatred have become progressively more significant over the last twenty years because of the growing presence of the internet in our professional, social, and private lives. The extensive use of social networking sites such as Twitter and Facebook, as well as our mounting utilization of online discussion forums and comments sections in newspapers, have led to a vast increase in the amount of ways in which cyberhate can now be perpetrated.

It is important at the outset to start with some important definitions. Cyberhate will be taken to mean the use of electronic communications to express in written form hateful comments, insults, or discriminatory remarks about a person or group of persons based on, for example, their race, religion, ethnicity, sexual orientation, disability, or transgender identity. This will incorporate all written content, including images such as memes, but will not include any forms of Voice over IP or instant messaging. It is important to state that 'cyberhate' is distinct from broader conceptions of 'hate crime', although it should be noted that certain forms of cyberhate might also be classified as a type of hate crime.

This chapter will consider the issue of cyberhate from the point of view of legal regulation. It will explore the particular challenges that globalization poses to the combating of online hatred effectively, and it will determine what the global response to cyberhate should be. First it will discuss the ways in which the legal regulation of hate speech online differs from the regulation of 'offline' hate speech. Then it will reflect on how globalization poses additional challenges to legal responses to online hatred. The third section will evaluate the current international response to cyberhate and will conclude with some suggestions for ways forward. The chapter argues that the issues surrounding freedom of speech and the harm caused by cyberhate need to be reconsidered in light of the way in which online hatred is committed. It will also suggest that a united global response is the most effective way to progress, while acknowledging that differing cultural and legal norms make this a slow and difficult process.

ISSUES OF PRINCIPLE

The first issue of vital importance is the need to justify, from a principled point of view, the criminalization of online hate speech. This is because if a state or international body proposes the creation of a new piece of legislation or legal framework, it needs to be demonstrated that there is a strong case in favour of using the law to prohibit such behaviour. Traditionally, J.S. Mill's (1991) notion of 'harm' has been used as a guiding principle, which stipulates that the criminal law should only be used to prohibit behaviour if it can be shown to cause 'harm' to others. Mill's 'harm' principle has been subjected to a number of criticisms and modifications over the years, but has had enduring appeal for many contemporary legal philosophers (e.g., Raz 1986). By contrast, legal moralists believe that a necessary (although not sufficient) condition for criminalization is that the underlying behaviour which the law aims to prohibit is morally 'wrong' (e.g., Finnis 1980). Within both schools of thought, views differ as to the essential qualities of 'harm' or 'wrongfulness'. For the purposes of a discussion on cyberhate, what this debate demonstrates is the need to articulate clearly the underlying 'harm' or 'wrong' which a particular piece of legislation seeks to outlaw.[1]

Identifying the harm caused by hate speech is also important in that it will enable states to ensure that any legal response constitutes a legitimate incursion into free speech. The issue of free speech is central to the debate on cyberhate. Free speech lies at the core of our democratic society for a number of reasons and is regarded as protection against abuse of power. J.S. Mill (1991: 20) in *On Liberty* wrote about the need for democratic states to ensure 'the fullest liberty of professing and discussing … any doctrine'. Such a right exists in most modern democratic states albeit in different

[1] In order to simplify the discussion, the term 'harm' will be used from now on whilst acknowledging that legal moralists would prefer the word 'wrong'.

forms such as for example the First Amendment under the US Constitution or Article 10 of the European Convention on Human Rights (ECHR). However, the right to freedom of expression is not absolute. J.S. Mill himself acknowledged that '… even opinions lose their immunity, when the circumstances in which they are expressed are such as to constitute … a positive instigation to some mischievous act' (1991: 56). As such, the right to freedom of expression can be restricted under certain conditions. Regulating hate speech is a prima facie violation of freedom of expression, and there has been a wide-ranging debate amongst hate crime scholars about the extent to which hate crimes are compatible with free speech (e.g., Blazak 2011) and a similar discussion has been undertaken by scholars on the relationship between hate speech and freedom of speech (e.g., Waldron 2012). The most commonly used argument is that the regulation of hate speech can be justified if to do so is warranted by the level of harm caused by the speech. Thus, this is an additional reason as to why it is imperative at the outset of a discussion of the regulation of cyberhate to have a clear understanding of the harm caused by online hate speech.

To some extent, the discussion about the harm in online hate speech will mirror the ongoing debate that hate crime scholars engage in when attempting to identify why hate crime or hate speech should be prohibited. Two particular explanations stand out as they have been very influential in the development of hate crime legislation. The first is that the harm lies in the fact that victims of hate crime 'hurt more' (Iganski 2001) and/or hate crime offenders are more blameworthy because of their bias motivation (Lawrence 2002). The second is that the harm in hate speech lies in the damage to public order or in the fact that it 'pollutes' the sense of security in the public space we all inhabit (Waldron 2012). Conceptualizing the harm of hate crime and hate speech along these lines results in two sets of distinctions being made: the first is between the *personal* and the *impersonal*, and the second is between the *public* and the *private*. If the harm in hate crime lies in the greater harm caused to the individual because of the offender's motivation, then hate crime legislation needs to reflect this. This results in legislation that treats hate crime as a problem that is *personal* to the victim, rather than an impersonal harm which is directed in a more abstract sense at a group of people. Meanwhile, if the harm in hate speech lies in the pollution of the public space or in the threat to *public* security and peace, legislation will focus primarily on protecting the public arena and will not be concerned with the exclusively private domain.

In some jurisdictions, these distinctions are intrinsic to the legal framework on hate speech. For example, in England and Wales the existing legislation which can be used in cases involving cyberhate can be divided broadly along the personal/impersonal divide and the public/private divide. On the one hand there are the offences under the Malicious Communications Act 1988, the Communications Act 2003, and the Protection from Harassment Act 1997 which are aimed at protecting the individual from behaviour that has caused the recipient or was meant to cause them harassment, alarm, distress, annoyance, inconvenience, or anxiety, such as an email sent to the victim threatening to assault them because they are Muslim. These offences are designed to deal with targeted attacks on individuals and are not aimed at impersonal attacks. On the other hand, the offences under the Public Order Act 1986 make it an offence to stir up hatred on the grounds of race, religion, or sexual orientation. An example that would fall under these provisions would be displaying in a public place a poster encouraging the killing of gay people. These provisions are aimed at regulating speech which is deemed to be a threat to public disorder as s 18(2) stipulates that these measures do not apply to speech that is expressed in a dwelling, or to speech that is only seen or heard by someone in a dwelling. Thus, the Act makes a distinction between words expressed in a home, and those expressed outside a home. The speech does not have to be aimed at anyone personally, and can be an impersonal attack on a group of people. This pattern can be seen in legislation in other countries as well. For instance, in the US there is a raft of state legislation which can give protection to individual victims of hate such as where threats have been made to the victim *personally*, or the behaviour is covered by stalking and harassment charges (Citron 2014: 123–5). Meanwhile the Supreme Court allows for the existence of hate speech laws as long as they only cover ‘fighting words’ (*Chaplinsky v New Hampshire*) which threaten *public* security by inciting imminent lawlessness (*Brandenburg v Ohio*).

Undoubtedly some online hate speech will fall into the category of either personal attacks or public incitement, and so would potentially be covered by the relevant legislation. However, not all hate speech appearing on the internet will fall neatly into this divide. This is because the nature of the internet and the way it is used challenges the traditional divisions between personal/impersonal and public/private. Thus, there is a need for the law to reconceptualize the notion of the harm caused by cyberhate in order to inform legal reform in this area, both at an international level and at a local level.

The distinction made between public and private discourse by the law is fundamentally challenged by the way in which the internet functions. Whilst some operations are clearly private—such as emails—what constitutes a 'public' space on the internet is harder to identify clearly. As already noted, under English and Welsh law, activities which take place in the 'home' are not covered by the public order offences (Public Order Act, s 18(2)). In other countries such as Canada a 'private conversation' is exempt from the provisions (s 319(2) of the Canadian Criminal Code). Nevertheless, at a very basic level, these distinctions between a private/public space will not work for the internet given that online hate speech can be written by someone in their home, and read by someone else in their home on the other side of the globe. Similarly, identifying what constitutes a 'private conversation' on the internet is not straightforward as different levels of privacy are expected depending on which internet service is being used. For example, whether a conversation over Twitter or Facebook or in private chatrooms is public or private may depend on the number of people involved, as well as the privacy settings adopted by different users. We may find it easy to argue that an online newspaper available to all without subscription is public, but there may be much disagreement about whether or not this is also true of all freely available websites, including blogs. In an era when there are serious concerns over government surveillance of the internet, it is problematic to suggest that all content available on the internet should fall into the public domain.

From a legal point of view, this requires some thought. Jeremy Waldron in his book *The Harm in Hate Speech* (2012) sets out some persuasive arguments in favour of prohibiting hate speech. He likens the harm in hate speech to pollution that contaminates the environment and poisons the atmosphere. Hate speech, he argues, undermines the dignity of those who are targeted, and weakens their ability to be treated as equals. For this reason, he argues, prohibiting hate speech is a legitimate incursion on freedom of expression. To illustrate this, he gives the example of a Muslim father who walks his children to school, and on the way he is confronted by posters with Islamophobic slogans, such as 'Muslims and 9/11! Don't serve them, don't speak to them, and don't let them in' (Waldron 2012: 1). These posters make him and his children feel unwelcome and threatened. This is a powerful image and demonstrates well how hate speech can deeply affect the daily lives of citizens. However, in order to employ these arguments to justify infringing free speech by regulating cyberhate, a further explanation is required because of some crucial differences between the online and offline world.

To begin with, as already argued, the 'public' space as defined in the physical world is more limited in scope and so, therefore, it is easier to defend placing restrictions on free speech. It is much harder to rationalize such curbs on our freedom when it involves our online behaviour which is now integral to our everyday lives. In fact, research suggests that young people see little difference in their online and offline identities, often making no distinction between the two (Miller 2013). Another difference lies in the way in which we come across material on the internet and the fact that we have varying degrees of control over what we encounter online. Whilst it might be possible to avoid looking at racist websites by not searching for them, or by not clicking on the relevant link if you accidentally come across them whilst searching for something else, there will be some situations where a user comes across hateful content whilst using a website that they have a reasonable expectation will not publish this sort of material, such as for example on the comments sections of an online newspaper. Interestingly, researchers have found that contrary to expectations, even on social media websites such as Facebook, users come across a substantial amount of ideologically diverse material (Bakshy et al. 2014). This suggests that we have less control over the content we encounter on social media than we would expect. Consequently, it might be necessary to consider a more nuanced approach to online hatred whereby differing levels of responsibility exist depending on the context. For instance, we might wish to impose an obligation on certain websites to police their online content, but might not want to necessarily extend this responsibility to *all* websites for fear that this would infringe free speech.

These fundamental differences between the online and offline world have profound implications for the debate surrounding the regulation of the internet. It means that we cannot simply rely on the same arguments used by scholars in favour of banning hate speech, such as those employed by Waldron. It also means that we have to go much further in giving particular thought to how we use the internet, how hate speech online is viewed and by whom, and consequently where its harm lies. A salutary illustration of the importance of constructing clear arguments in relation to free speech is the recent repeal of s 13 of the Canadian Human Rights Act. This made it an offence to communicate via the internet any 'undertaking . . . that is likely to expose a person or persons to hatred or contempt'. In its original form, s 13 applied only to telecommunications, and the infringement on free speech had been deemed legitimate and constitutional by the Supreme Court of Canada in *Canada (Human Rights Commission) v Taylor*. However, after s 13 was extended to undertakings over the internet in 2001, this initiated a very public debate about the constitutionality of s 13. Concerns were expressed, particularly over the freedom of the press (Walker 2013), and ultimately the pro-free speech arguments were deemed to be stronger and s 13 was finally repealed in 2013 (c. 37, s 2).

The impersonal/personal distinction that is usual in hate crime legislation might also not work as well when applied to hateful content on the internet. Whilst attacks on an individual by email or targeted attacks on Twitter could fall foul of offences which deal with injury caused to the victim, there may well be other ways in which hate perpetrated over the internet has far-reaching consequences that extend beyond harm to a particular individual. For example, Perry and Olsson (2009) have argued that the internet provides those who belong to groups we might broadly define as ones peddling 'hate' with the opportunity to 'retrench and reinvent … as a viable collective'

(<IBT>Perry and Olsson 2009: 185</IBT>). It allows them to establish a collective identity and, Perry and Olsson argue, could potentially lead to a ‘global racist subculture’ (ibid). The permanency of the material that appears on the internet also contributes to this. As these hateful messages can be read by anyone at any time, this means that hate on the internet transcends the victim/perpetrator model, and instead can help create a wider global hate environment. It could, therefore, be said that the harm in hate speech over the internet is not limited to the injury it causes to an individual but also lies in the damage it causes to the social fabric of our global society by bolstering and intensifying certain hate movements. This aspect of cyberhate is not currently given enough attention by the law.

Another characteristic of internet use that has important implications for cyberhate legislation is the dichotomy between on the one hand the ease with which comments can be published on the internet, and on the other the permanent nature of internet publication along with its global reach. This presents a difficult dilemma for legislators. The casual nature of much internet ‘chatter’ presents a challenge to legal regulation as these types of comments may lack the intention usually necessary for hate speech. This has led commentators such as Rowbottom (2012) to caution against the over-criminalization of low-level online speech. Notwithstanding, even if an intention to cause injury may be lacking in many such cases, the harm caused can nevertheless be considerable (Fearn 2014). It is also a paradox of internet use that even though the reach of the internet is much more extensive than traditional forms of print media, the regulation, whether legal or non-legal, is minimal in comparison. Newspaper editors, for example, will take carefully considered decisions about what material to include in their newspaper. Meanwhile, the internet makes it possible for anyone to easily disseminate to the public their views and ideas through a number of means such as the creation of blogs, by posting comments at the end of newspaper articles, or via social media. This means that those thoughts and ideas are not necessarily well-thought out, and can be intemperate and uncontrolled. This lack of restraint or self-regulation means that speech that in the past would not have been published by traditional print media is now available to all. The internet can also embolden those who would never threaten or attack another person in the physical world to do so online. This has led to a worrying increase in the number of people, particularly women, who have been subjected to online abuse, harassment, or threats. The European Agency for Fundamental Rights (FRA) found in a gender-based violence against women survey that one in ten women in the European Union (EU) said they had been the victim of cyber harassment since the age of 15, and it was found that 'cyber harassment' was the most common form of harassment that women had experienced in the twelve-month period running up to the survey (<IBT>FRA 2013</IBT>). It also found that young women were most at risk of being targeted in this way (ibid). FRA's survey of discrimination and hate crime against Jews also found that 10 per cent of Jews said they had experienced offensive or threatening anti-Semitic comments made about them online (ibid). Meanwhile, the American Association of University Women surveyed almost 2,000 students in 2010–11 and found that 30 per cent of them said they had experienced some form of online harassment (Citron 2014: 16).

This discussion highlights the fact that relying on traditional measures used to combat hate crime is not sufficient, and there is a need for specific legislation aimed at targeting digital hate crime (see also Guichard 2009). A large part of the legislative exercise will require a reconsideration of free speech in light of the points made above about the challenges of cyberhate. In the context of 'offline' hate crime and hate speech, this debate has been characterized by interplay between two important values: on the one hand there is the important democratic value of freedom of speech, and on the other there is the equally important concept of equality. Hate crime legislation is seen as an important tool in the push towards equality, but this necessarily entails an incursion on our freedom of expression. Citron (2014) and Brennan (2009) have both argued that when considering the extent to which an infringement on freedom of speech is justified by legal regulation of cyberhate, more emphasis should be placed on the concept of equality. Citron (2014) has explored cyberhate in depth and has focused particularly on the way in which women are exposed to hate online, and how this affects their ability to participate fully in society and thereby threatens the principle of equality. She outlines a number of examples of cyberhate attacks where women have abandoned their online presence as a consequence. Citron views this as an affront to civil liberties as these attacks hinder women from enjoying their right to access to education and employment. This illustrates how in the context of the discussion of online regulation, it is wrong to assume that the relationship between free speech and equality is mutually exclusive whereby any step towards greater equality necessarily involves an incursion into free speech. This is because in many instances, online regulation of hateful messages may in fact help preserve free speech rather than diminish it. If the internet is meant to be a platform for free speech, then this has to apply equally to all groups in society. If cyberhate is impeding this, then legal regulation must be employed to even the scales.

TECHNICAL AND PRACTICAL ISSUES

Having looked at some of the issues of principle pertaining to online hate crime, it is necessary next to consider some of the technical and practical concerns regarding the legislative response to cyberhate. These relate in particular to the international nature of the internet given that globalization poses a number of difficulties to the legal regulation of cyberhate. These challenges are of particular importance as they indicate that ordinary criminal offences such as those relating to harassment or stalking are not sufficient to deal with online hatred, and that what is needed instead is a legal response which is more tailored to the specific requirements of cyberhate.

The first challenge relates to the global nature of cyberhate. Unlike most ‘ordinary’ hate crime which is perpetrated by offenders who live in the same jurisdiction as the victims, online hatred can be carried out by offenders who live in a different country or even a different continent to their victim. This transnational nature of online hate crime generates a number of problems from a legal perspective. First, it would be necessary to decide which jurisdiction has authority to punish the perpetrator’s actions. From the victim’s point of view, there are disadvantages whatever the outcome. If jurisdiction lies with the perpetrator’s legal system, the victim will encounter a number of potential hurdles when attempting to access justice in a foreign country. There may be no existing legislation in the perpetrator’s legal system that covers the offending behaviour, and even if there is, there may be obvious cultural, linguistic, logistical, and financial difficulties which would confront the victim who is attempting to persuade the police in another country to take their case forward. If jurisdiction lies with the victim's legal system, justice can only be achieved if the perpetrator is extradited to the victim's country. Inevitably, this will not be straightforward as extradition treaties are neither universal nor comprehensive, and so there is no guarantee that a victim will find it easy to bring a perpetrator to court. This illustrates how cooperation between states is required. The international harmonization of laws relating to cyberhate, as well as effective extradition treaties, would go some way to solving these difficulties.

In practice, though, this is likely to be very difficult to achieve as different cultural and legal traditions across the world act as an enormous stumbling block to harmonization. For example, the US approach to free speech is very different to that of the ECHR. The US First Amendment right to free speech is framed in absolutist terms. This means that although in practice exceptions to the First Amendment do exist (Citron 2014), the rhetoric surrounding free speech is that of assertive guardianship. By contrast, freedom of expression under Article 10 of the ECHR is given protection subject to certain limitations. This has meant that the US has traditionally been less open to persuasion when drafting laws that may infringe freedom of speech. The divergence in approach to free speech extends beyond the US and Europe. Countries at different stages of political development will also have distinct needs in relation to the balance that needs to be struck between free speech and cyberhate. Whilst it could be argued that that the political development of Western liberal democracies allows for greater restrictions on free speech (Waldron 2012), other emerging nations might still require the scales to be balanced more in favour of free speech (Gagliardone et al. 2014). Given the importance of international cooperation and the crucial role the US would play in this, international agreement on cyberhate legislation appears a long way off (Banks 2011).

One particular practical difficulty associated with the regulation of cyberhate relates to the anonymity of the internet. Even if we were able to achieve international harmonization of laws and broker far-reaching extradition treaties, these will only be effective if it is possible to identify and track down the perpetrators. This is not necessarily straightforward given the ease with which it is possible to conceal one's identity online. Email accounts can be obtained using false information, and software is available that allows a user to hide the origin of an email or their physical location. Once a false email has been obtained, a perpetrator is able to join discussion groups, chat forums, and social media without their true identity becoming known to the victim. In countries where Internet Service Providers (ISPs) do not provide anonymous accounts it is possible, in principle, to obtain the Internet Protocol (IP) addresses of a perpetrator's computer, and thus track down the physical location of the computer which was used to carry out the attack. However, ISPs do not give out IP addresses easily and will often only do so under a court order. In the US, the courts have allowed 'John Doe subpoenas'[2] to be issued to a website or an ISP in order to compel the handing over of details of a particular poster in cases involving civil claims (Citron 2014: 223). Currently there is no similar provision in relation to criminal claims, and so these subpoenas are of limited use in most cyberhate cases which are better tackled through the criminal law.

A further problem is the issue of policing. It is imperative that any regulatory regime can be policed effectively if we are to justify the use of the criminal law and criminal punishment. Packer (1968) argues that one of the pre-conditions for the legitimate use of the criminal law is if doing so will not expose that process to severe qualitative and quantitative strains. This is a particular challenge for the policing of cyberhate given the scale and complexity of the internet. It is important to note, though, that the predominant role of the police with regard to cyberhate is to respond to complaints by individual victims, and not to police the entire World Wide Web. Thus, the police need to be given the appropriate tools and resources to deal effectively with individual incidences of cyberhate that are brought to their attention, but they do not need to be given the power or responsibility to regulate all online content. Coliandris (2012) suggests that a Problem-Oriented Policing (POP) approach would be necessary so as to identify the root cause of the behaviour before finding effective measures to deal with it. Clearly, there is still much work to be done to identify and

[2] These subpoenas enable a plaintiff to file suit even if they do not yet know the identity of the defendant.

develop best practice in this area and this will require a deep dialogue between governments, the security services, the police, and academics.

A combination of the global reach of the internet, the problems regarding policing, the difficulties surrounding international agreement over free speech, and the complications posed by anonymity, forces us to consider whether the responsibility for online hatred needs to be shifted to, or at least shared by ISPs. The current situation means that many victims of online hatred face considerable difficulties in locating their assailant, and even greater problems with bringing them to justice. In such cases, where tracing an offender is too complicated either because of geographical impediments or because of technical difficulties in identifying them, an alternative solution for the victim would be to force ISPs to take responsibility for policing their websites. This already happens in cases involving defamation in English law. Under s 5 of the Defamation Act 2013, website operators are given a defence to defamation claims if they comply with certain procedures. Thus, if they are able to show that they provided the complainant with sufficient details of the poster's identity which would enable them to bring proceedings against them, or they took down the offending material when asked to by the complainant, they will have a defence in law under s 5(3) of the Defamation Act 2013. Although there is no direct legal obligation on websites to provide identification information, or to take down offending speech, there is a very strong incentive for them to do so in cases involving defamation. It is too soon to evaluate the effectiveness of s 5, and its success depends on how clear the accompanying regulations prove to be. Arguably it is not an excessively onerous responsibility on ISPs, but the issue does raise broader questions about the importance of privacy on the internet. At a time when tensions are running high over state surveillance of internet communications both in the United States and elsewhere, any obligation on ISPs to hand over personal information to the authorities will have to be carefully worded in order to ensure this does not breach privacy rights. The main concern has been with unlimited and indiscriminate surveillance (e.g., UN 2013) and so it would be possible for similar criminal provisions to be drafted narrowly enough to strike a balance between freedom of speech and privacy rights on the one hand, and the right of victims of cyberhate to achieve justice by resolving some of the difficulties surrounding anonymity on the other.

Another important part of the solution to cyberhate will be a continuation and intensification of the informal techniques that already take place. Currently, there is a proliferation of different approaches that attempt to help solve the problem of online hatred in an informal non-legal way. For instance, the Anti-Defamation League (ADL) has produced a set of best practices which guide providers and the internet community on how best to tackle online hatred (ADL 2014). The emphasis is on cooperation by industry providers such as through the voluntary enforcement of terms of service and the provision of effective mechanisms for reporting and removing offensive comments (ADL 2015). There is a concern that over-reliance on private contractors to deal with cyberhate essentially gives those companies the power to make decisions on free speech, but equally, as a recent report by UNESCO on online hatred points out 'focusing exclusively on repressive measures can miss the complexity of a phenomenon that is still poorly understood and which calls out for tailored and coordinated responses from a range of different actors in society' (Gagliardone et al. 2015: 53). Other initiatives, such as Belgium's 'Stop Hate' website aims to give parents, teenagers, and teachers information and support on how to recognize and combat online hatred. The Council of Europe's 'Young People Combating Hate Speech Online' campaign has also sought to mobilize young people from around Europe to take a proactive stance against hate speech they confront online.

The globalization of online hatred means that in order to achieve the most effective response, international cooperation is required. However, this is not an easy solution as attempting to reach a consistent approach across different countries with vastly different legal systems and cultural approaches to issues such as free speech has proven to be very difficult. This will become clear in the following section which evaluates the attempts made so far at international collaboration. Nevertheless, the globalization of cyberhate also requires us to consider refocusing attention on those who enable it, rather than adopting a traditional response which concentrates on the perpetrators of hate. This corroborates the argument put forward in the first section above that we need to design a specific response to online hatred that requires us to think beyond the traditional measures usually adopted to combat physical hate crimes.

CURRENT INTERNATIONAL RESPONSE

At the global level, there has been a concerted effort over a number of years to establish a cohesive international response to hate crime. Article 20 of the International Covenant on Civil and Political Rights (ICCPR) and Article 4 of the International Convention on the Elimination of All Forms of Racial Discrimination both call on states to outlaw any dissemination or advocacy which incites discrimination or violence. The provisions in both articles are couched widely enough that legislation outlawing hate speech online (subject to freedom of expression) could be included. However, online hatred is not specifically mentioned, and so no guidance is provided on how these prohibitions would operate in relation to cyberhate. At the EU level, the EU Framework Decision on combating certain forms and expressions of racism and xenophobia by means of the criminal law requires states to harmonize their laws on racism and xenophobia. Whilst all of these initiatives constitute a constructive approach to hate crime, they do not provide a solution to any of the particular problems relating to online hatred outlined in the previous two sections. By the same token, the Organization for Security and Co-operation in Europe (OSCE) has produced under its Office for Democratic Institutions and Human Rights (ODIHR) a Practical Guide (2009) which provides a very useful framework for hate crime legislation, but it is less concerned with hate speech. Its recommendations will only apply to online hatred which falls into the category of an already existing criminal offence and, as such, they are of limited use to a wider discussion on cyberhate which seeks to establish both a reconceptualization of traditional notions of harm and a practical solution to some of the particular problems associated with digital hate speech. Thus, the current approach which subsumes cyberhate within the broader problem of hate crime is ultimately insufficient.

The only international agreement targeted expressly at online hate is the Council of Europe's Additional Protocol to the Convention on Cybercrime concerning the criminalization of acts of a racist and xenophobic nature committed through computer systems. To date, the Protocol has had limited success due to the reluctance by a number of states, such as the US, Russia, Turkey, Sweden, and the UK to ratify it. Brennan (2009) has written in detail about the shortcomings of the Protocol and she has argued that, even if ratified by a state, it is too narrow a document to provide an effective response to online hatred. She identifies a number of defects with the Protocol such as the reference to 'intention' which potentially limits the extent to which ISPs can be held liable for publishing hate material online, and the free speech exception under Article 3(3) which allows states to not implement cyberhate measures even if they ratify the Protocol. Furthermore, she contends that the emphasis on free speech undermines the underlying objectives of the Protocol, and she argues that more weight should instead be placed on the concept of equality and non-discrimination which states are under a duty to promote.

The legislative response to hate crime at the international level, does, therefore, appear to be insufficient. Nevertheless, there is a growing recognition that cyberhate should be treated as a distinct category of hate crime which requires special attention. For example, the report by the EU into the implementation of its Framework Decision points out that: '[d]ue to its special character, including the difficulty of identifying the authors of illegal online content and removing such content, hate speech on the internet creates special demands on law enforcement and judicial authorities in terms of expertise, resources and the need for cross-border cooperation' (European Commission 2014: para. 4).

Meanwhile, the working group on cyberhate at the EU Agency FRA conference on Hate Crime in 2013 considered the need for a targeted legislative approach to cyberhate (FRA 2013). In parallel to this, civil society groups have already taken steps to combat cyberhate as a discrete problem. The ADL treats cyberhate as a specific policy area, and has implemented a number of initiatives to combat online hatred such as the Cyber-Safety Action Guide which aims to provide internet users with a simple way of registering any hatred they find online. The International Network against Cyberhate (INACH) aims to bring the internet in line with human rights by uniting organizations to tackle online hate and to raise awareness of discrimination which takes place online.

CONCLUSION

Taken together, several themes emerge from the current international and national approach to hate crime. The first is that we need a targeted approach to cyberhate. There are a number of ways in which the nature and impact of cyberhate differs from offline hate speech, and which therefore require us to go beyond traditional definitions of harm to capture the complex and subtle distinctiveness of the damage caused by online hatred. Particular attention needs to be paid to the fact that online interactions between individuals can bolster hate movements, whilst the prevalence of online hatred on heavily used websites, such as online newspapers, may impact certain groups and affect their ability to enjoy the internet. In addition to this, more thought needs to be given to adopting a more nuanced approach to the concept of what constitutes a 'public' space on the internet. Finally, there will need to be a recognition that freedom of speech arguments need to be reconsidered in light of the harm caused by online hatred. In particular, the principle of equality as a competing value must be given greater attention in order to determine the correct limits of legal regulation.

The second point to make is that we need a globalized approach to cyberhate. The very nature of internet use means that the perpetrator and the victim will not necessarily be in the same country, and so without an international response, victims will often find they have no legal solution to their problem. In addition to this, we need to hold ISPs at least partially responsible for enabling cyberhate that comes to their attention, and this cannot easily be achieved without an international response. This international response also needs to take into account the unique nature of cyberhate. It is not sufficient to rely on efforts that are already being made at the international level in relation to hate crime more generally. Cyberhate needs to become a priority policy in its own right. Currently, the Council of Europe's Additional Protocol is the only international framework on cyberhate. Other international bodies such as the UN and the EU should reflect on the special requirements of online hatred as outlined in the previous paragraph, and establish agreements which are tailored towards the broader harm caused by cyberhate, and the need for social equality.

Third, although an international response is the ideal, it seems unlikely that a coherent approach will be achieved in the short-term. In the meantime, though, there is still much that can be done at the national level. National governments can still achieve results by adopting effective legislation aimed at online hatred. It is also imperative that European states which have not already ratified the Council of Europe's Additional Protocol do so because, in spite of its deficiencies, it signals a first step in the right direction.

The regulation of online hate speech is still in its infancy, and a number of challenges lie ahead. The difficulties are undoubtedly substantial, but the problems caused by cyberhate are equally significant and should not be underestimated. The regulation of cyberhate needs to become a top priority for those with responsibility in this field as the situation is pressing and action can no longer be deferred.

References

ADL. 2014. 'Cyberhate Responses. Best Practices for Responding to Cyberhate'. ADL website at: http://www.adl.org/combating-hate/cyber-safety/best-practices/#.VX_9vflViko

ADL. 2015. 'Report of the Anti-Defamation League on Confronting Cyberhate'. ADL website at: http://www.adl.org/assets/pdf/combating-hate/ICCA-report-2015-With-hyperlinks-May-8-2015_final.pdf

Bakshy, Eytan, Solomon Messing, and Adamic Lada. 2015. 'Exposure to Diverse Information on Facebook'. Facebook blogpost at: https://research.facebook.com/blog/1393382804322065/exposure-to-diverse-information-on-facebook

Banks, James. 2011. 'European Regulation of Cross-Border Hate Speech in Cyberspace: the Limits of Legislation'. European Journal of Crime, Criminal Law and Criminal Justice 19(1), 1.

Blazak, Randy. 2011. 'Isn't Every Crime a Hate Crime? The Case for Hate Crime Laws'. Sociology Compass 5(4), 244.

Brandenburg v. Ohio, 395 U.S. 444 (1969).

Brennan, Fernne. 2009. 'Legislating Against Internet Race Hate'. Information & Communications Technology Law 18, 123.

Canada (Human Rights Commission) v. Taylor [1990] 3 S.C.R. 892.

Chaplinsky v. New Hampshire 315 U.S. 568 (1942).

Citron, Danielle Keats. 2014. Hate Crimes in Cyberspace.

Coliandris, Geoff. 2012. 'Hate in a Cyber Age'. In Policing Cyber Hate, Cyber Threats and Cyber Terrorism, edited by Imran Awan and Brian Blakemore.

European Commission. 2014. 'Report from the Commission to the European Parliament and the Council on the implementation of Council Framework Decision 2008/913/JHA on combating certain forms and expressions of racism and xenophobia by means of criminal law'. At: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52014DC0027

Fearn, Harriet. 2014. 'Experiences of Online Hate Crime'. International Network of Hate Crime Scholars Conference.

Finnis, John. 1980. Natural Law and Natural Rights.

FRA. 2013. 'Working Group II. Challenges of Cyberhate'. At: http://fra.europa.eu/sites/default/files/frc2013-12-11-wg02-challenges_of_cyberhate.pdf

Gagliardone, Iginio, Alisha Patel, and Matti Pohjonen. 2014. 'Mapping and Analysing Hate Speech Online: Opportunities and Challenges for Ethiopia'. Working paper at: http://pcmlp.socleg.ox.ac.uk/wp-content/uploads/2014/12/Ethiopia-hate-speech.pdf

Gagliardone, Iginio, Danit Gal, Thiago Alves, and Gabriela Martinez. 2015. 'Countering Online Hate Speech'. Unesco Publishing at: http://unesdoc.unesco.org/images/0023/002332/233231e.pdf

Guichard, Audrey. 2009. 'Hate Crime in Cyberspace: The Challenges of Substantive Criminal Law'. Information and Communications Technology Law 18(2), 201.

Iganski, Paul. (ed.). 2009. Hate Crimes.

Iganski, Paul. 2001. 'Hate Crimes Hurt More'. American Behavioural Scientist 45(4), 626.

Irish Examiner. 2011. 'Man cleared of online hatred against Travellers'. At: http://www.irishexaminer.com/ireland/man-cleared-of-online-hatred-against-travellers-169325.html

Lawrence, Frederick. 1999. Punishing Hate: Bias Crimes under American Law.

Mill, J.S. (1991). On Liberty and Other Essays.

Miller, Danny. 2013. 'Future Identities: Changing identities in the UK—the next 10 years DR 2: What is the relationship between identities that people construct, express and consume online and those offline?' Government Office for Science. At: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/275750/13-504-relationship-between-identities-online-and-offline.pdf

Packer, Herbert. 1968. The Limits of the Criminal Sanction.

Perry, Barbara and Patrik Olsson. 2009. ‘Cyberhate: the Globalization of Hate’. Information and Communications Technology Law 18(2), 185.

Raz, Joseph. 1986. The Morality of Freedom.

Rowbottom, Jacob. 2012. ‘To Rant, Vent and Converse: Protecting Low Level Digital Speech’. Cambridge Law Journal 71(2), 355.

UN. 2013. ‘UN resolution to the right to privacy, a first step—UN expert on freedom of expression’. UN website at: http://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=14033&

Waldron, Jeremy. 2012. The Harm in Hate Speech.

Walker, Julian. 2013. ‘Canadian Anti-Hate Laws and Freedom of Expression’. Parliament of Canada Research Publications at: http://www.parl.gc.ca/content/lop/researchpublications/2010-31-e.htm#a8