Regulating hate speech online

BANKS, James

Published version

BANKS, James (2010) Regulating hate speech online. International Review of Law, Computers and Technology, 24 (3). 233-239. ISSN 1360-0869

Regulating Hate Speech Online

May 2010

Dr James Banks
Department of Criminology and Community Justice
Faculty of Development and Society
Sheffield Hallam University
Whitehouse
Southbourne
Collegiate Crescent Campus
Sheffield
S10 2BP

0114 225 5362 or 07805156121

j.banks@shu.ac.uk

James Banks is a Lecturer in Criminology and an Associate Researcher at the Hallam Centre for Community Justice, Sheffield Hallam University.

The exponential growth in the Internet as a means of communication has been emulated by an increase in far-right and extremist web sites and hate based activity in cyberspace. The anonymity and mobility afforded by the Internet has made harassment and expressions of hate effortless in a landscape that is abstract and beyond the realms of traditional law enforcement. This paper examines the complexities of regulating hate speech on the Internet through legal and technological frameworks. It explores the limitations of unilateral national content legislation and the difficulties inherent in multilateral efforts to regulate the Internet. The paper develops to consider how technological innovations can restrict the harm caused by hate speech whilst states seek to find common ground upon which to harmonise their approach to regulation. Further, it argues that a broad coalition of government, business and citizenry is likely to be most effective in reducing the harm caused by hate speech.

Key words: Hate Speech, Internet, First Amendment, Additional Protocol to the Convention on Cybercrime.

Introduction

As a globalised, decentralised and interactive computer network, the Internet was heralded by first generation Internet critics for its ability to cross borders, destroy distance and break down real world barriers.[1] Envisaged as an egalitarian communications medium, many commentators advocate a technological landscape unfettered by governmental regulation. Such a libertarian ethos stresses the transnational and borderless nature of the Internet, questioning both the legitimacy and ability of states to govern cyberspace. Cyber-libertarians also support the fundamental right of freedom of expression, arguing against the regulation and censorship of Internet content which could obstruct the free flow of knowledge, ideas and information.

Yet the anonymity, immediacy and global nature of the Internet has also made it an ideal tool for extremists and hatemongers to promote hate. Alongside the globalisation of technology, there has been an incremental rise in the number of online hate groups and hate related activities taking place in cyberspace. Unsurprisingly:

> As computers become less expensive, simpler to use and consequently more common in ... homes (and workplaces), as the barriers to disseminating information through computers fall, bigots of all kinds are rushing to use the power of modern technology to spread their propaganda.[2]

Whilst the itinerant nature of hate sites makes accurate quantification virtually impossible, the general consensus amongst monitoring organisations and scholars is that the number of sites has grown exponentially. The first extremist website, Stormfront.org, was launched by white nationalist and former Ku Klux Klan member Donald Black in April 1995. By the turn of the century, it was recorded that the number of bigoted websites had soared to around 400.[3] The most recent estimate, from the Simon Wiesenthal Center, suggests that there are

currently around 8,000 hate sites in existence.[4] Social networking sites such as *Facebook* and *Myspace* have also become breeding ground for racists and far-right extremist groups to spread their propaganda. Accessed by more than 200 million users, such websites, which attract thousands of new members every day, have become a key conduit through which extremists can educate others, transmit ideas and beliefs, and mobilise for demonstrations and rallies.

The Internet has become the 'new frontier'[5] for spreading hate, as millions can be reached through an inexpensive and unencumbered social network that has enabled previously diverse and fragmented groups to connect, engendering a collective identity and sense of community. Perry and Olsson suggest that previously isolated and atomised members of far-right groups have been replaced by a 'global racist subculture'[6] who share values, ideologies and fears.[7] Technological innovation has enabled extremists and hatemongers to propagate their rhetoric and strategies, recruit, organise and unify through websites, private message boards, listservs and email.

The growth in online hate groups has been mirrored by the rise in web-based hate speech, harassment, bullying and discrimination, targeted directly and indirectly through forums, blogs and emails. This rise in hate speech online is compounded by difficulties in policing such activities which sees the Internet remain largely unregulated. Criminal justice agencies are unlikely to proactively dedicate time and money to investigate offences that are not a significant public priority. Consequently, the police will rarely respond to online hate speech unless a specific crime is reported.

Yet despite such shortcomings, more and more nation states have sought to combat the publication of hate propaganda. This paper examines the complexities of regulating hate speech on the Internet through legal and technological frameworks. It explores the limitations of unilateral national content legislation and the difficulties inherent in multilateral efforts to

regulate the Internet. The paper develops to consider how technological innovations can limit the harm caused by hate speech whilst states seek to find common ground upon which to harmonise their approach to regulation. Further, it argues that a broad coalition of government, business and citizenry is likely to be most effective in limiting the harm caused by hate speech.

Unilateral Regulation of Hate Speech

States' criminalisation of the publication of hate propaganda has been followed by more recent efforts to prosecute individuals for the dissemination of racist and xenophobic material online. Despite the geographic indeterminacy of the Internet, political and technological developments have seen states seek to impose virtual borders onto cyberspace in order to regulate online hate speech. Yet unilateral efforts to legislate against offensive material have been stymied by limited jurisdictional reach and conflict that has occurred when states have sought to enforce laws extraterritorially into other jurisdictions. This is unsurprising given that national laws on hate diverge so widely, as a consequence of countries 'unique socio-political responses to unique problems'.[8] In particular, the US First Amendment affords considerable protection to those espousing hate from American websites, in direct contrast with many other nations approach to hate speech. The case of *Yahoo!* demonstrates the drift towards nation states imposing geographical demarcations onto the virtual world and more pertinently highlights the difficulties inherent in European countries seeking to extend their jurisdiction extraterritorially, enforcing their content laws against material uploaded beyond national boundaries.

In the landmark case of *Yahoo!, Inc v. La Ligue Contre Le Racisme et L'Antisemitisme,* two French student organisations sought to prosecute Internet Service Provider (ISP) *Yahoo!* for contravening a French law forbidding the offering for sale of Nazi

merchandise. Article R. 645-1 du Code Penal outlaws the wearing or public display of insignia, emblem or uniform of an organisation or individual responsible for crimes against humanity, as such behaviour is deemed to be a serious crime 'against the people, the state and public safety'.[9] It was alleged that *Yahoo!* had violated this law by displaying Nazi memorabilia on its auction website.

Although the content originated in the United States, the French court ruled that *Yahoo!* was liable and should seek to eliminate French citizens' access to the sale of Nazi merchandise. *Yahoo!* argued that its actions lay beyond French territorial jurisdiction, as the material was uploaded in the US where such conduct is protected by the First Amendment. Dismissing this claim, Judge Jean-Jacques Gomes applied an effects-based jurisdictional analysis and granted prescriptive jurisdiction, describing the sale of Nazi paraphernalia as an 'insult to the collective memory of a country profoundly wounded by the atrocities committed by the Nazi Enterprise'.[10] The court ruled that intentional transmission, in addition to the local impact of the visualisation of Nazi memorabilia, provided sufficient grounds for finding jurisdiction. Recognising that the application of geo-location technology, combined with a declaration of nationality from service users, would filter out 90 per cent of French citizens, the court ruled that *Yahoo!* apply such mechanisms in order to seek to reduce access to the sale of Nazi merchandise. Failure to comply with the court order within three months would result in *Yahoo!* becoming subject to a penalty of 100,000 francs per day.

Yahoo! subsequently sought and received, from the United States District Court for the Northern District of California, a judicial ruling that the enforcement of the French authorities would breach the First Amendment of the Constitution. The US court made it clear that whilst hate speech is odious, it will be protected under the First Amendment unless it can be demonstrated that such speech contains a direct, credible 'true' threat against an identifiable individual, organisation or institution; it meets the legal test for harassment; or it

constitutes incitement to imminent lawless action likely to occur.[11] District Judge Jeremy Fogel ruled that whilst the enforcement of foreign judgement is founded upon a "comity of nations", the court could not put into effect a policy which would undermine its own fundamental interests:

> The French order's content viewpoint-based regulation, whilst entitled to great deference as an articulation of French law, clearly would be inconsistent with the First Amendment if mandated by a court in the United States. What makes this case uniquely challenging is that the Internet in effect allows one to speak in more than one place at a time. Although France has the sovereign right to regulate what speech is permissible in France, this Court may not enforce a foreign order that violates the protections of the United States Constitution by chilling protected speech that occurs simultaneously within our borders. [12]

The judicial impasse of the *Yahoo!* case exemplifies the cultural tension inherent in attempts to regulate online speech extraterritorially. Whilst nation states are able to successfully prosecute hate crime that takes place within their own territorial boundaries, they have not been able to extend their reach beyond their borders. Consequently, online hate speech which originates in one jurisdiction, but whose effects are felt elsewhere, continues to go unregulated. As Kaplan, Moss, Lieberman and Wessler recognise, the protection afforded by the Internet means that 'a perpetrator of a threat or harassing speech need not be at the actual scene of the crime (or within 5,000 miles, for that matter) to prey on his or her victim.'[13] With states unable to overcome the obstacles of the Internet's anonymity and multijurisdictionality, hate crime offenders are free to thwart the somewhat piecemeal and arbitrary assemblage of national laws. In light of such inadequacies, the Council of Europe

introduced a protocol aimed at harmonising national legal system's computer related offences in order to reach a common minimum standard of relevant offences and enable cooperation in the prosecution of those committing hate crimes in cyberspace. Discussion now turns to assess the efficacy of this provision.

Multilateral Regulation of Hate Speech

With unilateral attempts to regulate hate speech originating in foreign territories falling foul to jurisdictional and cultural conflict, the application of national law to foreign entities has serious limitations. Consequently, an international system governed by compacts and supranational decision making would appear to offer an appropriate means through which to obviate regulatory conflict between nation states. However, such a collaborative enterprise has been seriously undermined by the US's commitment to free speech. So whilst the US approach to regulation has become a minority view, its indirect unilateralism detracts from European efforts to construct a truly international regulatory system.

The Council of Europe's Convention on Cybercrime is the first multilateral compact that seeks to tackle computer based crime by increasing the cooperation amongst nations, harmonising national laws and investigatory techniques. Although the United States is not a member of the Council of Europe they do have observer status. Non-European countries may also be invited to sign and ratify council treaties in order to broaden the scope and impact of agreements. The US has both signed and ratified the Convention on Cybercrime which provides a multilateral framework for tackling a variety of Internet crimes, including child pornography, copyright infringement and fraud. However, the US's signature was only obtained after an Internet hate speech protocol, which was initially included in the Convention on Cybercrime, was removed at their behest.

In response, the Council of Europe introduced a separate protocol to address hate speech online. Under the Additional Protocol to the Convention on Cybercrime, parties are required to criminalise acts of a racist and xenophobic nature committed through computer systems. By the start of 2010, the additional protocol had been signed by 32 member states and ratified by 15. The US has, however, informed the Council of Europe that it will not be party to the protocol as it is inconsistent with their constitutional guarantees. So whilst the European compact has undoubtedly increased coordination and cooperation in combating hate crime, the US's commitment to indirect unilateralism is extremely problematic. This is because most hate sites originate in the US and although many target an American audience, leakage is inevitable.

Alongside criminalisation, the protocol extends the scope of the Convention's extradition provision to include those sought for Internet hate speech crimes. Yet the US has no bilateral extradition treaties with European countries and therefore no commitment to deliver defendants to be charged with committing hate speech offences. Consequently, European countries are left in a double bind being both unable to enforce civil judgements in American courts and unable to extradite American offenders for criminal prosecution. This is likely to result in the US increasingly becoming a safe haven for propagators of hate speech:

> Given that the US with our First Amendment essentially is a safe-haven for virtually all Web content, removing content or shutting down a Web Site in Europe or Canada through legal channels is far from a guarantee that the contents have been censored for all time. The borderless nature of the Internet means that, like chasing cockroaches, squashing one does not solve the problem when there are many more waiting behind the walls – or across the border. Many see prosecution of Internet speech in one country as a futile gesture when the speech

> can re-appear on the Internet, almost instantaneously, hosted by an ISP in the United States.[14]

Whilst the Additional Protocol is a laudatory endeavour, it is undoubtedly limited in its ability to bring together real differences in the ways in which states envisage hate speech and construct a legal framework through which hate based conduct may be counteracted. In particular, European nations' commitment to combating the growth of online hate is undermined by the US First Amendment which provides a refuge for many of those propagating hate. Increased cooperation between European countries may simply result in an increase in the number of bigoted websites originating in the US as racists and extremists seek to avoid prosecution by moving their operations to America. Accordingly, transference rather than prevention is the likely outcome of increased legal regulation by the European community.

Technological Regulation of Hate Speech

With clear difficulties inherent in unilateral and multilateral legislative approaches to regulation, it is fair to question whether the law alone is the most appropriate means through which to counteract cyberhate. Recourse to technological regulation at both user and server ends may offer an effective avenue through which to minimise the transmission and reception of online hate speech.

Internet Service Providers can play a crucial role in reducing the level of online hate available to Internet users. Codes of conduct or Terms of Service (TOS) agreements enable ISPs to remove offensive web material that breach their policies. Voluntary codes of conduct to which customers must consent offer an important mechanism through which to regulate

websites originating in the US as they circumvent the First Amendment. This enables ISPs to delete content or cancel their service if TOS agreements are broken.

Some ISPs have actively sought to regulate hate speech by removing offensive content. For example *America Online* removed the neo-Nazi Website, the Nationalist Online, from its server for violating its terms of service agreement, which prohibits content that is racially or ethnically offensive. Furthermore, in countries such as the UK, ISPs have formed industry associations, which enforce codes of conduct prohibiting hate speech. In the UK, these codes have been produced despite The Electronic Commerce Directive (Hatred against Persons on Religious Grounds or the Grounds of Sexual Orientation) Regulation 2010 absolving ISPs and other digital service providers from liability for hate speech communicated across their networks.

Yet problematically, the vast majority of the many thousands of access providers in the US do not regulate against hate speech per se. Many TOS agreements are extremely narrow in focus, so whilst libellous or defamatory speech is prohibited, such codes of conduct do not extend to those acts that fall within the First Amendment's free speech protections. This is because Section 230 of the Communications Decency Act specifically states that ISPs will not be held criminally responsible for the speech of their users. Accordingly, there is little motivation for ISPs to self regulate. Curiously, many companies defend their hosting of hate sites through recourse to the First Amendment despite the fact that they are not bound by its protections.

Nevertheless, monitoring organisations, such as the Anti-Defamation League (ADL) and Simon Wiesenthal Center, continue to work closely with ISPs in identifying and removing hate based websites and messages and materials that contravene TOS contracts. ADL also seek to highlight material that is harmful in content and have been particularly successful in providing the public with information concerning online hate materials. In 2004,

Google responded to a ADL's concerns about access to anti-Semitic Website 'Jew Watch' by introducing an 'offensive search results' link that explains why some extremist websites appear in users search findings. More recently, *YouTube* has introduced an ADL authored guide on tackling hate speech in its Abuse and Safety Centre.

Undoubtedly, there are serious limitations to ISP based technological regulation. With removal contingent on commerciality and cost, many ISPs do not have the desire or means through which to address harmful content. Alternatively, governments can seek to block extraterritorial websites that do not comply with national laws. Geographic location technology can further enable both servers and states to control the flow of information on the Internet. Geo-location tools can identify the users IP address and, in turn, their location, in order to restrict access and filter out odious material.

Web users can also employ software, such as firewalls, to filter out sites containing certain speech. Numerous commercial Internet filtering software packages are readily available and can easily be installed on home computers. In 1998, ADL introduced Hatefilter, a filtering software product that not only prevents access to websites that promote hate, but also educates users about the nature of bigotry and why such sites should be rejected.

Individual responses to online hate may only have a limited impact on access to online material, but the responsibilisation of individual users can both promote a culture of intolerance towards online hate and contribute to efforts to 'reclaim' the web. Users can also play an important role in monitoring Internet content and alerting relevant authorities to incidents of cyberhate which may warrant law enforcement intervention. Hate speech hotlines have become an effective means through which citizens can alert ISPs to material that breaks their codes of conduct or law enforcement agencies to sites or incidents of cyberhate which may warrant investigation and prosecution.

As Bailey neatly summarises, 'broad-based efforts involving strategic alliances among citizens, citizen coalitions, industry and government provide a strong foundation from which to engage in visible, publicly accountable action against cyberhate.'[15] For such an alliance to operate effectively, governments, businesses and citizenry must all engage in individual and collective solutions to minimising online hate speech.

Conclusion

The exponential growth in the Internet as a means of communication has been emulated by an increase in far-right and extremist web sites and hate based activity in cyberspace. The anonymity and mobility afforded by the Internet has made harassment and expressions of hate effortless in a landscape that is abstract and beyond the realms of traditional law enforcement. States have sought to regulate the domain of the Internet through the conventional strategy of national law. However, the multi-jurisdictionality of the Internet has undermined states efforts to place geographical demarcations onto cyberspace. Furthermore, European efforts to harmonise national laws have been undermined by the US's commitment to the First Amendment, which provides a substantial safe haven for many of those transmitting hate.

Unilateral and multilateral efforts to regulate online through criminal law alone will not be enough to reduce the effects of online hate. The episodic prosecution of individual web users is unlikely to deter others from posting hate speech online. Web sites that are closed in one jurisdiction may simply re-open in another thus remaining available to Internet users worldwide. Furthermore, the global nature of the Internet makes the total legal regulation of cyberspace impossible. Consequently, it is necessary to seek alternatives through which to both limit the publication of hate speech online and minimise the harm caused by such behaviour. By combining legal intervention with technological regulatory mechanisms –

monitoring, IPS user agreements, user end software and hotlines – the harm caused by online hate can be diminished. Moreover, through the careful integration of law, technology, education and guidance, a reduction in the dissemination and impact of online hate speech can be achieved without adversely affecting the free flow of knowledge, ideas and information online.

---

[1] D. R. Johnson and D. Post, 'Law and Borders: The Rise of Law in Cyberspace', Stanford Law Review, 48 (1996), 1367-1402; J. R. Reidenberg, 'Governing Networks and Rule-making in Cyberspace', Emory Law Journal, 45 (1996), 911-929; J. T. Delacourt, 'The International Impact of Internet Regulation', Harvard International Law Journal, 38 (1997), 207-235.
[2] Anti-Defamation League, Hi-tech hate: Extremist use of the Internet (New York: ADL, 1997).
[3] B. Levin, 'Cyberhate: A Legal and Historical Analysis of Extremists' Use of Computer Networks in America', American Behavioural Scientist, 45, no.6 (2002), 958-988.
[4] Simon Wiesenthal Center, Digital Terrorism and Hate 2.0 (Los Angeles: SWC, 2008).
[5] Anti-Defamation League, Combating Extremism in Cyberspace: The Legal Issues Affecting Internet Hate Speech (New York: ADL, 2000).
[6] L. Back., M. Keith and J. Solomos, 'Racism on the Internet: Mapping neo-fascist subcultures in space', in Nation and Race, ed. J. Kaplan and T. Bjorgo (Boston: Northeastern University Press, 1998), 73-101.
[7] B. Perry and P. Olsson, 'Cyberhate: the globalization of hate', Information and Communications Technology and Law, 18, no. 2 (2009), 185-199.
[8] Harris et al. suggest that each country's hate speech laws are rooted in historical and philosophical and constitutional traditions in relation to freedom of expression. C. Harris., J. Rowbotham, K. Stevenson, 'Truth, law and hate in the virtual marketplace of ideas: perspectives on the regulation of Internet content' Information and Communications Technology and Law, 18, no. 2 (2009), 155-184.
[9] *La Ligue Contre La Racisme et L'Antisemitisme (LICRA) and Union Des Etudiants Juifs De France (UEJF) v. Yahoo! Inc. and Yahoo France.* English translation available at http://www.juriscom.net/txt/jurisfr/cti/yauctio ns20000522.htm (5 February 2010).
[10] Ibid.
[11] C. D. Van Blarcum, 'Internet Hate Speech: The European Framework and the Emerging American Haven', in Computer Crime, ed. I. Carr (Aldershot: Ashgate, 2009), 327-376.
[12] *Yahoo! Inc. V. La Ligue Contre Le Racisme et L'Antisemitisme*, 169 F. Supp. 2d at 1192("The extent to which the United States, or any state, honours the judicial decrees of foreign nations is a matter of choice, governed by 'the comity of nations.'" (quoting Hilton v. Guyot, 159 U.S. 113, 163 (1895))).
[13] J. E. Kapplan., M. P. Moss., M. L. Lieberman and S. Wessler, Investigating hate crimes on the Internet (Washington: Partners Against Hate, 2003).
[14] C. Wolf, 'Hate speech on the Internet and the law', available at http:adl.org/osce/osce_legal_analysis.pdf (5 February 2010).
[15] J. Bailey, 'Strategic Alliances: The inter-related roles of citizens, industry and government in combating Internet hate, Canadian Issues, Spring (2006), 56-59.