Social Media, Violence and the Law: "Objectionable Material" and the Changing Contours of Hate Speech Regulation in India

By Siddharth Narrain

Abstract

With the advent of the internet and increasing circulation of hate speech, and material that has been linked to public order disturbances, there has been a shift in the legal discourse around hate speech. What has emerged, especially post the striking down of section 66A of the Information Technology Act, are categories such as 'objectionable', 'provocative' content. The focus has shifted from the content itself, what it says, and the intention of the author, to being able to pre-empt the circulation of such material. Law is increasingly invoked to prevent speech (through prior restraint) rather than post facto investigation and prosecutions.

This in turn has given rise to a range of institutional mechanisms such as monitoring labs that are now part of policing practice. Additionally, civil society organizations are now collaborating with police to help trigger mechanisms to take content off internet platforms. Increasingly it is through keywords and algorithmic searches that the category of hate speech has been defined rather than traditional legal doctrine. In the words of Lawrence Lessig, code plays the role of law, and the architecture of the internet becomes policy.

This paper will examine the issues outlined above relying heavily on a series of interviews with lawyers, policy analysts, journalists, academics, civil society activists, and police personnel conducted in Delhi, Bengaluru, Mumbai and Pune.

Introduction

Debates around the regulation of hate speech are highly contested globally. There is little agreement over what constitutes hate speech, what part of hate speech should be regulated by law, and where to draw the line between freedom of speech and hate speech that is deemed illegal. However, the range of international, domestic and theoretical material that has emerged around this theme, helps us understand and situate hate speech and the impulse to legally define and regulate such speech.

This paper traces the legal and regulatory debates that have emerged in response to hate speech in India, situating these developments within the global context. It focuses on the shift from regulation of content online to the regulation of circulation of content, as evident from specific examples of incidents of inter-group violence in Bengaluru (Bangalore), Pune (Poona), and Western Uttar Pradesh during the last decade. By examining these specific instances where content online, including morphed images of gods and goddesses, attacks on persons revered by Scheduled Castes and Tribes, and rumours meant to incite violence and exacerbate communal tensions, this paper argues that the increasing use of internet-enabled mobile phones and peer-to-peer communication platforms could potentially lead to a fundamental shift in the manner in which law and governance mechanisms respond to inter-group violence.

I have relied largely on detailed interviews conducted between January 2015 and March 2017 with police, lawyers, civil society groups, journalists, academics and policy experts. I have also relied on news reports from this period, legal documents such as First Information Reports, judgments and orders, and a range of secondary material including published books and articles. While I have focused almost entirely on India, I refer to the United States and European context in the introduction to broadly situate the Indian context within global developments around the theme of regulation of hate speech.

Globally, countries have taken very different approaches to regulating hate speech. One of the outliers in terms of hate speech law is the United States, whose Supreme Court has set a high bar for what kind of speech can be construed illegal, and the First Amendment has been interpreted widely to provide a robust mechanism for protection of speech. This approach remains substantially the same when it comes to online hate speech. This liberal approach to hate speech of United States law has been criticized by Matsuda and other critical legal scholars who have argued that United States law does not account for the way in which for example, racist speech perpetuates historical inequalities and harms communities that have been at the receiving end of targeted violence (Matsuda et al 1993). The philosopher Jeremy Waldron, in his influential book, The Harm in Hate Speech Law, argues that the standards set by the United States Supreme Court do not account

for instances where hate speech targeting or vilifying a group can lead to a feeling of insecurity among the target group and prevent those from that group from participating freely in the public sphere (Waldron 2012).

In contrast to the United States, European nations have traditionally placed more restrictions on hate speech. The German government has been one of the most pro-active in the world in regulating hate speech online. Faced with increasing anti-immigrant speech in platforms Facebook and Twitter, has enacted The Act to Improve Enforcement on the Law in Social Networks (NetzDG) enacted in October 2017 that is aimed at ensuring that social media platforms regulate content that is already illegal under the German Criminal Code. This law requires that social media platforms set up effective and transparent complaint mechanisms for the regulation of hate speech and other online content that is illegal, and has been severely criticized by German opposition parties as curbing free speech (Theil 2018).

The way the German law is formulated indicates a tension between what is typically be considered public communication in Western liberal democracies, when compared to parts of the Global South including countries like India. Since this law specifically exempts email and messaging apps it is not clear whether it would be effective when it comes to peer-to-peer networks such as WhatsApp, which is very widely used in India. The picture becomes more complicated when we consider that in countries like India and Indonesia powerful political parties (Chaturvedi 2016) and other religious and political interest groups (George 2016), have actively used social media to amplify such content to further their own political or social agenda.

Internationally, legal principles around freedom of speech evolved through the Universal Declaration on Human Rights (UDHR), the International Covenant for the Protection of Civil and Political Rights (ICCPR), the European Convention on Human Rights (ECHR), and the Inter-American Commission on Human Rights (IACHR) allow for regulating speech that incites hatred or discrimination against a group. However, such a restriction must be provided by law, must not be arbitrary, and must pursue one of the legitimate aims listed – such as protecting the rights and reputation of others, or to protect national security, public order or the rights or public health and morals. The laws that are enacted to restrict speech should be both necessary and legitimate to achieve these aims, and should also use least restrictive proportional means to achieve their purported aim (Electronic Frontier Foundation et al 2014: 14). These principles are meant to ensure that there are checks and balances built into this kind of regulation of speech.

Scholars have used categories such as 'extreme speech' and 'dangerous speech' to demarcate a category of speech that is seen as generally unacceptable. Hare and Weinstein (2009), for instance, define extreme speech as a form of speech that is

outside acceptable norms of dissent. Anushka Singh, in a recent book, *Sedition in Liberal Democracies,* describes extreme speech as a wider category of speech that liberal democracies find uncomfortable including categories such as hate speech, sedition, pornography, and libel. (Singh 2018: 3-14). Johnathan Maynard and Susan Benesch use the category of 'dangerous speech' to demarcate speech that is unacceptable. Benesch draws upon her experience of working with international tribunals that have dealt with the question of genocide. In Rwanda, for instance, calls for violence over the radio played a crucial role in laying the groundwork for the violence that followed. Maynard and Benesch describe dangerous speech a category of speech that has a reasonable chance of catalyzing or amplifying violence by one group against another, keeping in mind the circumstances in which it is being disseminated and factors such as who is speaking, the nature of the audience, and the means of communication (Maynard and Benesch 2016). For the purposes of this paper I will use the term hate speech as a broader category that includes both extreme speech and dangerous speech.

In India, the term 'hate speech law' is used by lawyers and the media to refer to a number of laws that proscribe 'promoting enmity' between classes of people, (e.g. section 153A of the Indian Penal Code (IPC)) and outraging religious sentiments (e.g. section 295A of the IPC). Hate speech laws have their origin in colonial British policy, especially in the deep seated view that law makers such as Thomas Macaulay had that Indian subjects were especially vulnerable to insult and offence and were religiously and emotionally excitable subjects, quick to create public disorder on provocation based on insult to religion or religious beliefs (Ahmed 2009: 173). While hate speech laws have evolved gradually over time through judicial pronouncements (Narrain 2016), the use of social and digital media to spread hate has brought into focus specific legal and governance issues. Increasingly there is recognition that the posting and circulating of hate speech online poses specific challenges linked to the speed, scale, and volume of transmission across multiple platforms and formats, which can be interlinked (Gagliardone et al 2015: 13).

Hate speech laws in India are medium neutral – they apply equally to theatre, print, radio, broadcasting, and the internet. However, each time a media technology has gained popularity, special laws have been enacted to govern these, which are usually based on the language of the penal provisions governing hate speech. In case of the internet, a 2008 amendment to the Information Technology Act, 2000 (hereafter, IT Act), passed with barely any discussion in Parliament, resulted in the enactment of, amongst other provisions, section 66A of the IT Act that criminalized offensive content– defined as content that caused "annoyance", "insult", "enmity", "hatred or ill will" etc. Initially meant to tackle spam, this provision came to be used frequently against content online that was thought to have the capacity to cause public order disturbances. In 2015, section 66A was struck down

by the Indian Supreme Court as violative of the right to freedom of speech and expression. Even while 66A was in operation, the police often used 66A along with the Indian Penal Code (IPC) provisions such as Section 153A, that prohibits enmity between groups and committing acts prejudicial to maintenance of harmony, and section 295A that proscribes speech that outrages religious feelings or insults religious feelings or beliefs.

The Central Government has the power to block content online under section 69A of the IT Act. The most common ground for invoking section 69A is disturbance of public order (Arun et al 2018: 136). The procedure to be followed to block content is contained in the Information Technology (Procedure and Safeguards for Blocking of Access of Information by Public) Rules, 2009 (or Blocking Rules, for short). Section 79 of the IT Act governs the liability of intermediaries for content, exempting them from liability subject to certain conditions (Ibid).

Thus in India hate speech online is governed by a combination of colonial era penal laws and laws specifically enacted to regulate communication online. All of these are subject to 'reasonable restrictions' to the Article 19(1)(a) of the Indian Constitution, which guarantees the freedom of speech and expression. These include 'public order', which is the most commonly invoked restriction. When compared to other jurisdictions, Indian law regulating hate speech is closer to the European model, and is far more restrictive than in the United States.

'Objectionable Material', Communal Violence and Social Media

The last decade has seen a gradual rise in the widespread use of social media in India, enabled by the availability of affordable smartphones, affordable data plans, increased broadband penetration, and the expansion of internet use in regional and vernacular languages. The current decade in India has seen the emergence of a media discourse that links 'social media' to 'public order disturbances'. In many of these incidents, it is hate speech provisions that have been regularly invoked, along with other provisions dealing with public order. Dana Boyd refers to social media as 1) sites and services that emerged globally in the early 2000s, including social network sites, video sharing sites, blogging and micro blogging platforms, and related tools that allow participants to share their own content and 2) the cultural mindset that emerged in the mid 2000s as part of the technical and business phenomenon called Web 2.0. (Boyd 2014:6). For the purposes of this paper I use the term social media to include digital media such as MMSs, SMSs, and communication platforms such as WhatsApp, Telegram and Snap Chat, given that much of the content I am talking about moved between these communication platforms with great ease. I have consciously included peer-to-peer encrypted Over the Top

(OTT) platforms such as WhatsApp, as the volume of content on this platform is extremely high with over 200 million monthly active users in India alone (Financial Express 2018). Further, the ability of WhatsApp groups with a maximum of 256 members to effectively circulate and magnify content, and the ease with which text, audio, and video files can be shared on WhatsApp, and other similar OTT platforms, make them central to the circulation of hate speech online.

The Legal Regulation of 'Objectionable Material'

From 2010 onwards there have been a number of incidents of violence reported by the news media that have been linked to content that has circulated on social media. Such material is now popularly referred to in the news media as 'objectionable' material. Going by the dictionary meaning of 'objectionable', this would amount to material that arouses distaste or opposition, is unpleasant or offensive; it thus covers a wide range of material that is not just related to hate speech or incitement to violence.

In cases related to communal violence and hate speech, this term 'objectionable material' is used to both allude to and elides the exact nature of the content. This works well for news reports, as it avoids the problem of the content of these messages, by virtue of being reported, leading to further provocation or tensions in an already tense situation. 'Objectionable material' has over a period of time come to stand in for 'hate speech', 'seditious material', 'obscene material', and defamatory material. 'Objectionable material' has been used since the colonial period and continues to be used by governments and media, to refer to the larger category of material that the government has taken off the internet (or other media) because the material is not appropriate for viewing or consumption. For instance, the first legislative intervention to regulate cinema in India, the Indian Cinematograph Act, 1918, was justified by the colonial government as necessary to prevent the screening of objectionable films (Hughes 2000:51).

The most important development in Indian law related to internet is the 2015 Supreme Court decision in the Shreya Singhal case, where the young student Shreya Singhal challenged section 66A of the Information Technology Act (Shreya Singhal 2015). This case related to a challenge to three key provisions of the Information Technology Act 2000, the law that governs the internet in India. These provisions related to the proscribing of 'grossly offensive' and 'menacing' content (section 66A), the government's authority power to block content (section 69A) and the legal standards that governed internet intermediaries (section 79). The case garnered publicity in the backdrop of a series of arrests of artists, students, and those critical of political figures under section 66A. These arrests happened both at the central and regional level, and were severely criticized by civil society and human rights organizations, especially those working in the area of technolo-

gy, free speech, and digital rights.

In one of these incidents, two young women were arrested by the Maharashtra police for violating section 66A of the Information Technology Act by posting a comment on Facebook criticizing the fact that city of Mumbai (Bombay) shut down after the death of the Shiv Sena (a powerful local political party) leader Bal Thackeray. One of them had posted the comment and the other had 'liked' this post. The comment resulted in Shiv Sena members physically threatening the two women, and vandalizing a clinic owned by one of the two women's relatives. (Press Trust of India 2012).

The arrests prompted widespread outrage (Arun et al 2018: 134), and led to a young student Shreya Singhal challenging section 66A of the Information Technology Act (Ibid) in the Supreme Court. Other petitioners including civil society organizations and internet trade associations intervened in this case broadening the scope of the challenge to include sections 69A and 79.

In the Shreya Singhal case, the Supreme Court struck down section 66A of the IT Act but upheld section 69A (power to block content) and 79 (due diligence for intermediaries), quoting the Additional Solicitor General (representing the Central Government), who argued that there should be a different standard of reasonable restrictions for the internet. The Additional Solicitor General, while justifying this proposition stated that "the recipient of the free speech and expression used in a print media can only be literate persons while internet can be accessed by literate and illiterate both since one click is needed to download an objectionable post or a video" (emphasis added) (Ibid at Para 27). Thus the government attempted to justify the retention of a law meant to curb free expression online on the basis that there was something different about the medium, specifically pointing to the ease of use and the fact that one did not require to be literate or have specialized knowledge to use this medium. The democratization of the internet is used as a justification to ensure greater regulation of the medium, and to apply a different legal standard that would allow for the government to restrict free speech online.

The Northeast Exodus, 2012

One of the first incidents of inter-group violence, where hate speech was circulated on SMSs, MMSs and posts on Facebook was in 2012. It targeted persons from the Northeast of the country[1] living in cities of south and west India, such as Bengaluru, Pune and Chennai (Madras), and it subsequently led to a mass exodus. A large number of persons from the Northeast live in these cities because of better employment and educational opportunities, when compared to cities and towns in the Northeast. Persons from the Northeast are often the target of discrimination in these cities given their distinct ethnicities, and cultural practices

(McDuie-Ra 2017: 27-44).

In August 2012, tensions began to flare in Bengaluru after messages began circulating online threatening retribution for those from the Bodo community killed during the July 2012 ethnic violence in Kokrajhar, in the Northeastern state of Assam between indigenous Bodo community and Bengali-speaking Muslims, who have faced the brunt of anti-immigrant sentiment in the state. These messages explicitly threatened retribution by Muslims against persons from the Northeast in the lead up to Eid al-Fitr, a Muslim festival celebrated at the end of Ramadan (Ibid: 34).

Many of these SMSs, MMSs and Facebook posts were circulated widely by persons belonging to the Northeast to their friends and family, creating a ripple effect within the community. This led to an unprecedented feeling of insecurity among persons from the Northeast, as well as anyone who looked like they were from this region (which included Tibetans, Indian Chinese, and Koreans), leading to an estimated thirty thousand people fleeing these cities in a mass exodus within a span of few days (Sailo 2012). In Bengaluru, many who fled were students and labourers. Just how much the circulation of material on social media was linked to this feeling of insecurity can be gauged from the Central government's measure to ban bulk SMSs temporarily and to block websites (even though these were arbitrarily chosen), which was seen by many persons from the Northeast at the time as a necessary measure to address the fear and mass panic that was created by SMSs, MMSs and threatening messages circulating on social media.

This incident became one of the first recorded instances where large numbers of people, fearing for their security and safety left their places of residence, after receiving threatening messages on their phones through SMSs and MMSs. The Northeast exodus led to a lively debate within civil society and policy groups on whether the government's response in temporarily barring bulk SMSs and blocking a number of websites amounted to overreach (Prakash 2012), and just how crucial a new form of technology, mobile phone enabled digital media, was to such mass panic (Sundaram 2012).

Muzaffarnagar Riots, 2013

In 2013, the Western part of Uttar Pradesh, the most populous state in India, witnessed large-scale political violence against Muslims. Thousands of Muslims were displaced from their homes, and continue to live in makeshift homes in the districts of Shamli and Muzaffarnagar. One of the key factors in the violence was the active role of politicians including the BJP legislator from Uttar Pradesh, Sangeet Som, who posted a controversial video on his Facebook page. This controversial video, circulated on social media in the lead up to the violence in 2013, helping to mobilize people for a Jat mahapanchayat (a gathering of local village councils

of the dominant Jat caste grouping that is a traditionally agricultural community) community on 31 August 2013. The gathering was specifically called as a response to the ongoing violence in the region and the circulation of the video allowed for the mobilization of large numbers of persons from the community, which in turn acted as a catalyst to the large-scale violence that unfolded subsequently. At the time of the violence, the video was circulated as showing a Muslim mob lynching two Hindu men. The video was later debunked as being filmed in an unrelated incident in Sialkot, Pakistan in 2010 (Centre for Policy Analysis 2013). However at the time it was circulated, and in the context of communal tensions that already existed, the impact of the video should not be underestimated. Newspaper reports indicate that even the Commission of Inquiry into the Muzaffarnagar riots point to the failure of the state government to effectively respond to the circulation of this video, thus acknowledging the role of its circulation during the violence that ensued (Ali 2016).

Inter-group Violence in Pune, 2014

In this section I build on my earlier fieldwork in Pune where I had examined the role of the police and civil society in the aftermath of inter-group violence in 2014 (Narrain 2017). In order to address the problem of hate speech online during incidents of inter-group violence, the police and sections of civil society have, over a period of time, begun to realize how best to activate the internal mechanisms of intermediaries to remove "objectionable content". A striking example of how this was achieved occurred in Pune, in the aftermath of communal riots in events in the months leading up to June 2014. A group of individuals who called themselves The Social Peace Force, who had already worked in this area around the issue of drought relief in 2013 and had a Facebook Group of more than 20,000 members, decided to intervene to prevent such objectionable material from circulating online To do this they assumed the role of civil society watchdogs, and began to monitor content on social media. They used key words included terms such as "Ram", "Sita", "Laxman", (Hindu religious figures) and "Allah". These keyword searches were done in multiple spellings and pronunciations.

Once they detected material they considered unacceptable, they would send this to a group of 10 people amongst them whose role was to look at the content and decide if they should ask for it to be taken down. If they decided that the post was a problem, they would call upon the 20,000 strong Facebook group to report this content as spam to Facebook (Ibid). Faced with such large number of spam reports, Facebook began taking down such content immediately. Of course even in asking and responding to such material, the group had to be careful to not violate the existing hate speech law and section 66A of the IT Act (Ibid). Eventually they approached the police, and the then Minister of State for Home of Maharash-

tra, Satej Patil who advised them to involve the Pune Police's Cyber Cell. Soon members of Pune Police's Cyber Cell and the Minister joined the Social Peace Corps Facebook group. At the time their effort came in for severe criticism from a prominent writer who accused them of moral policing – a viewpoint more widely representing those uncomfortable with what they saw as a form of vigilantism online. The group now continues to cooperate with the Cyber Police to ensure that such material is removed (Ibid). The Social Peace Force identified what is objectionable through a 'non-discriminatory'/multi-faith model. A member of the Force who I interviewed said, "If a God's image is replaced with a model's body, we would identify this as bad. We did not discriminate based on religion".

What this shows is that the architecture of the internet and technological capacities of citizens and governments determine the manner in which objectionable material is regulated, rather than legislation and administrative orders. For instance, in Pune, the Social Peace Force suggested another method of controlling communal violence that has been implemented by the city police. On WhatsApp, the police created groups of police stations, housing societies, social workers, and politicians, with 3,300 police officers added to different groups. This helped them have a substantial presence on WhatsApp communities to track images and utterances so as to respond quickly. Thus they combined traditional mechanisms of police surveillance with usage of new technological platforms, and in partnership with civil society.

Police Surveillance, Pedagogy, and Publicity

Besides such local examples, over the last few years institutional mechanisms centered on surveillance, monitoring, and training and equipping police to deal with cyber crimes have emerged. The most prominent of these is the Mumbai Social Media Lab (MSML) set up in 2014. The MSML was set up in collaboration with the National Association of Software and Services Companies (NASSCOM) (a trade association of software companies) and The Data Security Council of India (a not for profit industry body set up by NASSCOM) using a monitoring application provided by a for profit company, SocialAppsHQ.Com, to help real time alerts on content related to social media platforms. The MSML was set up as an immediate response to the massive mobilization of protestors as well as public anger in December 2012 sparked off by the brutal gang rape of a young woman in Delhi. The Mumbai police publicly stated that the MSML would help them keep tabs on the "mood and emotions of citizens" and track public views and sentiments on "sensitive issues and protests" (Press Trust of India 2013).

The second set of Social Media Labs are being set up in the state of Uttar Pradesh, by the state police in collaboration with a public university, the Indraprastha Institute of Technology (IIIT), Delhi. While one lab is located in Meerut

the communally sensitive Western region of the state, the other is in the capital, Lucknow. Official statements in the media indicate that the main reason for this has been concern around communally sensitive material circulating in Uttar Pradesh, especially in the wake of the Muzaffarnagar riots of 2013 discussed earlier in this paper (Bhatia 2015).

Despite these public and highly publicized interventions by the Social Media Labs in Mumbai and Uttar Pradesh, the Union Telecommunications Minister did not mention of these in responding to a Parliament question in April 2015 on whether the government was monitoring social media sites. Instead, he stated that there is no institutional monitoring mechanism for monitoring social networking sites, and that Law Enforcement and Intelligence / Security Agencies monitor the internet on a case-to-case basis. The Minister only referred to the Electronic Media Monitoring Centre (EMMC), located in the Union Ministry of Information and Broadcasting, to track trends on social media and monitor "public interface on the social media network." (Ministry of Information and Technology 2015). Moreover, the Minister also referred to section 79 of the IT Act that requires intermediaries to observe due diligence quoting an advisory issued to all intermediaries by the previous government in 2012, to monitor "both national and international networking sites", and to disable inflammatory and hateful content hosted on their websites on a priority basis (Ibid).

Pedagogy and publicity have emerged as important instrument in the police's effort to deal with 'objectionable material'. For instance in the aftermath of the 2014 Pune riots, the police embarked on an extensive campaign to educate the public on the dangers of circulating material during communal riots. This happened both offline and online. They organized lectures and debates in educational institutions in Pune with the help of IT experts, teachers, and social activists. Large hoardings over the city urged the public to not "like" or "dislike" content that was communally sensitive, nor to post comments, share, or forward such material (Narrain 2017). The police organized meetings in public spaces, housing societies and community halls. Through these meetings, the police encouraged the public to report such content circulating online (Ibid).

Code as Law

The examples from the previous sections illustrate how police and civil society struggled with the question of hate speech online, seemingly caught unaware by the scale and speed of circulation and impact of such content. Over a period of time there has emerged a collaborative effort to respond in real time to the circulation of hate speech online, the most effective of these being in Pune, which has a strong technological infrastructure and persons who were well-equipped to adapt

to this challenge.

These strategies bring me back to Boyd's definition of social media, as both about its technological aspects as well as the cultural practices associated with it, which include the impulse to forward, like and share content (Boyd 2014). Boyd's emphasis on the link between the technological and cultural aspects of social media can be read in conjunction with the work of the legal scholar Lawrence Lessig. In his influential book 'Code 2.0', Lessig argued that the architecture of cyberspace becomes a de facto regulator, and the technology underlying the internet, or code can be compared to law (Lessig 2006: 78-79). Referring to chat rooms in the USA, for instance, Lessig says that the fact that only 23 people are allowed in an AOL (America Online) chat room is the choice of code engineers, but the effect of this is that it becomes much more difficult to excite members of AOL into public action. Lessig argues that, although AOL was one of the largest internet Service Providers in the world at that time, with 27 million in 2006, the architecture of the space only allowed for a maximum of 23 persons to gather in one space together. Lessig argued that on AOL, there was no space large enough for citizens to create a riot (Lessig 2006: 90-91). If one compares Lessig's AOL chatroom example with contemporary WhatsApp groups that allow for 256 members and are easier to use on a continuous basis because of internet enabled mobile phones, the role of code in Lessig's formulation becomes clearer.

Transposing this example to the situation around 'objectionable' material online, it seems as if there are many versions of law and policy at play. Along with the law laid down by sovereign states, we have guidelines and community standards that are formulated by global companies, such as Facebook, Google, and Twitter. In addition to this there are technical capacities of police to monitor and filter information, done primarily through key word searches. Then we have the ability of police and civil society to trigger mechanisms by intermediaries such Facebook flagging hate speech posts. These technical capacities determine the level and effectiveness of regulation of content, as much as   laws formulated by states. In the current scenario, the ability or inability of governments to intercept or block material on messaging platforms such as WhatsApp, Instagram, and Snapchat have led to both the ability to escape state regulation and enhance extreme measures by the Central and state governments. These extreme measures include suspending all internet services for extended periods of time, and blocking particular platforms such as Facebook, Twitter, and WhatsApp.

The Union Telecommunication Minister, in a reply to a parliament question in Winter 2015 revealed that during January and March 2015 the government had not only blocked (under section 69A, IT Act) 143 urls, including those of Facebook, Twitter, Orkut, and Linked In, but also asked social networking sites to disable 496 urls in order to comply with court orders (Lok Sabha Unstarred

Question 2015). These compliances are usually done through take down orders (under section 79 of the IT Act) or through self-reporting tools instituted by intermediaries. These self-reporting tools include mechanisms such as flagging on YouTube, where users can flag material if they violate YouTube's community guidelines. These guidelines are bunched under categories such as 'hateful content', 'violent and graphic content', 'harmful or dangerous content', 'nudity or sexual content', copyright violations and threats (YouTube Community Guidelines). As per Facebook's Governments Requests Report, it restricted 1228 pieces of content between January and June 2017. The majority of these were because of violating laws related to hate speech and defamation of religion (Facebook Data Requests January-June 2017).

These statistics indicate that the response from law and order and government to hate speech online has been to request Facebook to take down content, and to take the drastic measure of shutting down the internet. In the next section, this paper will examine the move from post publication prosecutions to prevention action by law enforcement mechanisms in more detail.

Conclusion: Prosecution to Prevention

Along with the emergence of these new mechanisms of governance to tackle objectionable material online, the focus of law enforcement has moved increasingly from responding after the publication to preventive action. In other words, focus is moving away from the content and the originator of the content to preventing the circulation of such content. Part of this is linked to technological difficulties in identifying whom to hold culpable, especially when material is downloaded and moves with consummate ease across platforms such as YouTube, Facebook and WhatsApp. Another factor that has resulted in this move is jurisdictional limitations when it comes to prosecuting material that originates outside the country. The Indian government has also been putting pressure on intermediaries to locate their servers within India to allow for greater access and control of such information by agencies of the government.

The most obvious example of this shift can be seen in the manner that the Union and state governments have resorted to internet bans in specific geographical locations. One of these bans, the Gujarat government's blocking of internet sites in a number of cities across the state in 2015, was challenged by a law student in the High Court of Gujarat. The government's decision to cut off access to mobile internet connections was in response to political agitations related to the demand for reservations in government jobs by the Patidar community, a dominant caste in the state of Gujarat. The agitation was far from peaceful, with incidents of violence and arson, and curfew imposed by the state government in several cities

and towns across the state. The state police had directed telecommunication companies to stop services in cities across Gujarat for over a week (Singh 2015).

The main ground of challenge in court was that the government used Section 144 of the Code of Criminal Procedure Code (CrPC) to block access to internet over mobiles, while it should have used its powers under section 69A of the IT Act to block select sites or pages. Section 144 CrPC is a law that has traditionally been used to enforce curfew and maintain law and order during or in anticipation of riots and public order disturbances (Gaurav Sureshbhai Vyas). The state government argued that the public order situation was so serious that they had to resort to the use of Section 144 of the CrPC, and that this section was targeted at persons rather than just internet sites. This justified the use of Section 144 CrPC instead of section 69A of the IT Act and related blocking rules, which are meant to block specific urls or websites.

The use of Section 144 CrPC signals the onset of legal responses meant to prevent internet access rather than prosecute persons based on content, or even block specific content based on filtering mechanisms. This mass scale shutting down of internet access was justified by the Gujarat High Court as meeting the standards under Article 19(1)(a) of the Indian Constitution which guarantees the fundamental right to freedom of speech and expression. The High Court, in its judgment, agreed with the state government's contention that since the shutdown extended only to mobile internet services in the state, and broadband and wifi access to internet was available, the government had applied its mind and not taken an arbitrary decision. The court expressed its faith in the executive's capacity to take a call on how best to respond to a public order situation, thus giving the government plenty of leeway in the means they used to restrict access to the internet (Supra Gaurav Sureshbhai Vyas). The constitutionality of Section 144 CrPC has been challenged earlier, but courts have so far upheld it, stating these measures are needed in urgent situations where there are public order disturbances. (Arun et al 2018: 142)

Using the example of hate speech law in India, specifically online content deemed 'objectionable' linked to public order disturbances, I have attempted to show that the broader questions around the governance of hate speech have shifted from a focus on the content itself to managing circulation of such content, which is mediated by new media technologies.

The sheer velocity, and reach of social media has changed the rules of the game. In effect the technological change in the period I describe in this paper, has led to a situation where law and regulation of 'objectionable material' has become more complex, and difficult for states. The movement of content across territorial borders, the importance of guidelines and rules created by transnational companies such as Facebook, Twitter and Google, monitoring and censorship

through algorithmic searches, new mechanisms of policing, and closer engagements between state and civil society organisations have together created a landscape that has shifted the contours of the law around hate speech. The perceived importance of dealing with 'objectionable material' within government and police has led to regulatory structures and practices, which while resonating earlier histories of regulating print, broadcasting, and cinema, have inaugurated a distinct moment.

Siddharth Narrain is a lawyer and legal researcher based in Delhi. He is a Visiting Research Fellow at the Sarai Programme, Centre for the Study of Developing Societies (CSDS), Delhi. E-mail: siddharth.narrain@gmail.com

Notes

[1] The category 'Northeast' refers to a geographical area in India that includes many communities and identities, many of who have been fighting for their right to self-determination. The expression was originally coined as a bureaucratic term, which arose during the reorganization of states in the 1960s and 1970s as a result of the border war with China. While this category was intended to homogenize and depoliticize, it is a category that people and media from this region have begun to use widely, especially with reference to their position within the Indian state and in relation to mainstream Indian society, Mcduie-Ra, Duncan, (2017): "Solidarity, Visibility, Vulnerability: 'Northeast as a Racial Category in India", Yasmin Saikia & Amit R. Baishya, Northeast India: A Place of Relations", New Delhi: Cambridge University Press.

References

Ahmed, Asad Ali (2009): "Specters of Macaulay: Blasphemy, the Indian Penal Code, and Pakistan's Postcolonial Predicament", Raminder Kaur and William Mazzarella (eds): *Censorship in South Asia: Cultural Regulation from Sedition to Seduction*, Indiana: Indiana University Press, 172-205.

Ali, Mohammad (2016): "A Strategic Commission of Enquiry", *The Hindu*, 11 March 2016, http://www.thehindu.com/opinion/columns/muzaffarnagar-riots-a-strategic-omission-of-inquiry/article8337424.ece (accessed on 03/03/18).

Arun, Chinmayi et al, (2018): *Hate Speech Laws in India*, New Delhi: Centre for Communication Governance, National Law University Delhi.

Bhatia, Ishita, (2015): "Meerut gets Second Social Media Lab in the Country after Meerut", *The Times of India*, Mumbai, 21 August, 2015, http://timesofindia.indiatimes.com/city/meerut/Meerut-gets-2nd-social-media-lab-in-country-after-Mumbai/articleshow/48594285.cms (accessed 15/06/17).

Boyd, Dana (2014): *It's Complicated: The Social Lives of Networked Teens*, New Haven:Yale University Press.

Centre for Policy Analysis, (2013): "Muzaffarnagar 2013: Violence by Political Design", 17 September 2013, http://sacw.net/article5612.html (accessed 10/06/17).

Chaturvedi, Swati (2016): *I am a Troll: Inside the Secret World of the BJP's Digital Army*, New Delhi: Juggernaut.

Electronic Frontier Foundation & Article 19, (2014): *Necessary and Proportionate: International Principles on the Application of Human Rights Law to Communications Surveillance* http://www.ohchr.org/Documents/Issues/Privacy/Electronic-FrontierFoundation.pdf (accessed on 01/03/2018).

Facebook Data Requests June to December 2017, https://transparency.facebook.com/ (accessed on 01-10-18).

Financial Express, (2018): "WhatsApp now has 1.5 Billion Active Users, 200 million Users in India", 1 February 2018, https://www.financialexpress.com/industry/technology/whatsapp-now-has-1-5-billion-monthly-active-users-200-million-users-in-india/1044468/ (accessed on 20/05/18).

Gagliardone, Danit, Gal, Danit, Alves, Thiago & Martinez, Gabriela, (2015): *Countering Online Speech*, Paris: United Nations Educational Scientific and Cultural Organisation: http://unesdoc.unesco.org/images/0023/002332/233231e.pdf (accessed on 02/03/2018).

Gaurav Sureshbhai Vyas v. The Petitioner who is a Law Student, C/WPPIL/191/2015, Gujarat High Court, 15 September 2015, https://indiankanoon.org/doc/29352399/, (accessed on 01/10/18).

George, Cherian, (2016): *Hate Spin: The Manufacture of Religious Offense and its Threat to Democracy*, Cambridge, MA: The MIT Press.

Hare, Ivan and Weinstein, James, (eds.) (2009): *Extreme Speech and Democracy*, Oxford, Oxford University Press.

Hughes, Stephen, (2000), "Policing Silent Film Exhibition in Colonial India", Ravi Vasudevan (ed), *Making Meaning in Indian Cinema*, New York: Oxford University Press.

Lessig, Lawrence (2006): *Code Version 2.0*, New York: Basic Books.

Matsuda, Mari J., Lawrence, Charles R., Delgado, Richard & Crenshaw, Kimberly Williams (1993): *Words that Wound: Critical Race Theory, Assaultative Speech and the First Amendment*, Boulder, CA: Westview.

Lok Sabha Unstarred Question No. 5939, Answered on 29 April 2015.

Maynard, Jonathan and Benesch, Susan (2016): "Dangerous Speech and Dangerous Ideology: An Integrated Model for Monitoring and Prevention", *Genocide Studies and Prevention: An International Journal*, 9:3, 70-95.

Mcduie-Ra, Duncan, (2017): "Solidarity, Visibility, Vulnerability: 'Northeast as a Racial Category in India", Yasmin Saikia & Amit R. Baishya, *Northeast India: A Place of Relations*, New Delhi: Cambridge University Press.

Ministry of Information and Technology, (India), Lok Sabha Unstarred Question No. 5939, Answered on 29 April 2015, "Objectionable Content on Web".

Narrain, Siddharth (2016): "Hate Speech, Hurt Sentiment and the (Im)Possibility of Free Speech", *Economic and Political Weekly*, 11:17, 119-126.

Narrain, Siddharth, (2017): "Dangerous Speech in Real Time: Social Media, Policing and Communal Violence", *Economic and Political Weekly Engage*, 24 August 2017: http://www.epw.in/engage/article/dangerous-speech-real-time-social-media-policing-and-communal-violence (accessed on 20/05/18).

Prakash, Pranesh, (2012): "Analysing Latest List of Blocked Sites (Communalism and Rioting Edition), 22 August 2012: https://cis-india.org/internet-governance/blog/analysing-blocked-sites-riots-communalism (accessed on 20/05/18).

Press Trust of India, (2012), "Two Mumbai Girls Arrested for Facebook Post against Bal Thackeray get Bail", IndiaToday.In, 19 November 2012, https://www.indiatoday.in/india/story/2-mumbai-girls-in-jail-for-tweet-against-bal-thackeray-121882-2012-11-19 (accessed on 06/11/18).

Press Trust of India, (2013): "Mumbai Police Tracks Social Media to Gauge Public Sentiment", *IndiaToday.In*, 14 July 2013, https://www.indiatoday.in/india/north/

story/mumbai-police-social-media-sensitive-issues-protests-170273-2013-07-14 (accessed on 20/05/18).

Sailo, Laldinkima, (2012): "The Great 'Exodus': Violence in Assam and its Aftermath", Institute of South Asian Studies National University of Singapore Insights, 6 September 2012: https://www.isas.nus.edu.sg/wp-content/uploads/media/isas_papers/ISAS%20Insights%20187%20-%20The%20Great%20Exodus.pdf (accessed 20/05/18).

Singh, Anushka, (2018): *Sedition in Liberal Democracies*, New Delhi: Oxford University Press.

Singh, Neha, (2015): "Gujarat Patel Agitation: Ban on Mobile Internet, Whatsapp Lifted in Ahmedabad?" *IB Times Television*, 1 September 2015.

Shreya Singhal & Ors v. Union of India & Ors, AIR 2015 SC 1523.

Sundaram, Ravi, (2012): "The Art of Rumour in the Age of Digital Reproduction", *The Hindu*, 19 August 2012: http://www.thehindu.com/news/national/the-art-of-rumour-in-the-age-of-digital-reproduction/article3792723.ece (accessed on 20/05/18).

Theil, Stefan (2018): "The New German Social Media Law: A Risk Worth Taking?" UK Human Rights Blog, 19 February 2018: https://ukhumanrightsblog.com/2018/02/19/the-new-german-social-media-law-a-risk-worth-taking-an-extended-look-by-stefan-theil/ (accessed on 20/05/18).

Waldron, Jeremy (2012): *The Harm in Hate Speech*, Cambridge, MA: Harvard University Press.

YouTube Community Guidelines, http://www.youtube.com/yt/policyandsafety/en-GB/communityguidelines.html (accessed 11/05/18).