Hate Speech and Social Media in Sri Lanka

Sanjana Hattotuwa & Roshini Wickremesinhe

In: *Acts of Media : Law and Media in Contemporary India*, Siddharth Narrain (edt.), pp.141-163

Series: Politics and Society in India and the Global South
Year of Publication: 2022
Publisher: Sage (New Delhi)
URL: https://perspectivia.net/publikationen/psigs/narrain_media/hattotuwa_wickremesinhe_media
Year of Publication on perspectivia.net: 2022

CHAPTER 7

HATE SPEECH AND SOCIAL MEDIA IN SRI LANKA

Sanjana Hattotuwa and Roshini Wickremesinhe

INTRODUCTION

With online communication now the norm amongst youth, Facebook, Twitter and instant messaging services such as WhatsApp have birthed a culture where individuals who share interests, views and opinions are able to interact with others beyond borders and continents, sans any social restrictions, forming, in some cases, movements and even new identity groups. Sri Lanka is no exception to this phenomenon. The power of such movements is not restricted to mere rhetoric and images displayed on a screen. The metamorphosis to real-time action or even violence fuelled by hate and fear—imagined or real—is a worrying and growing global concern (Gagliardone et al. 2015), more so since the catalysts of such movements and their followers are as young as teenagers. Garret Keizer writes, '… as if thirteen- and fourteen-year-olds who couldn't boil an egg are capable of creating a culture. They grow on what we feed them. It has never been otherwise. The only thing that changes is the food' (2014). That is the challenge: ensuring that what is served up by social media is wholesome.

THE EVOLUTION OF SOCIAL MEDIA IN SRI LANKA

The short message service (SMS) technology of more than a decade ago and e-mailing enabled instant messaging and sharing of information with wider groups. These were soon overtaken by newer social media technologies and mobile apps that afford constant Internet connection and real-time relaying of information. According to the Sri Lanka Department of Census and Statistics, in 2015, Internet usage is the highest among youth between the ages of 20–24 years (27%) and 25–29-year-olds are the second-highest users (25%; Department of Census and Statistics, Sri Lanka 2015). A survey by the Centre for Policy Alternatives on the 'Consumption of Mainstream and Social Media in the Western Province' indicates that 18–24-year-olds 'don't engage with mainstream media as newspapers, radio or TV, rather they engage with main stream media content through smart phones, Facebook and chat apps.' It also noted that 51 per cent preferred going online for news (Centre for Policy Alternatives 2016). Online social media and web-based platforms, accessed increasingly through smartphones and tablets, have emerged as an important and necessary vent for critical dissent in a context where the mainstream media does not and cannot afford the space for questioning or content that holds the government accountable for crimes and corruption. The growth of content creation and consumption online, wider and deeper than any other medium in the country and at an accelerated pace, has also resulted in low-risk, low-cost and high-impact online spaces to spread hate, harm and hurt against specific communities, individuals or ideas. As per Facebook's user statistics in 2016, Facebook, Twitter, as well as instant messaging for mobile services, such as Instagram and WhatsApp, are used, with Facebook emerging as the most widely used social network in Sri Lanka with 3.1 million users.

HATE SPEECH ON SOCIAL MEDIA

Conspiracy theorists, fringe lunatics and trolls have, since the first days of the Internet, inhabited online spaces and engaged with devoted followers or sought to deny and decry those who question them. The growth of hate speech can be seen as a natural

progression outward from these pockets of relative isolation and is also pegged to the economics of broadband Internet access and the double-digit growth of smartphones—an underlying, coast-to-coast network infrastructure capable of rich media content production and interactive, real-time engagement. This infrastructure has erased traditional geographies—hate and harm against a particular religion, identity group or community in one part of the world or country can, for example, within seconds, translate into violent emulation or strident opposition in another part, communicated via online social media and mediated through platforms such as Twitter, Facebook and instant messaging services.

The term 'hate speech' encompasses speech intended to intimidate, degrade and incite prejudice or even violence against a community based on their nationality, race, ethnicity, religion or sexual orientation. Online hate speech alone does not amount to hate crimes (physical expressions of violence) but often acts as a catalyst for physical manifestations of violence. A considerable amount of social media hate speech in Sri Lanka occurs on Facebook. Hence, many of the references in this chapter will address hate speech on Facebook. The ability to like, share and comment on posts allows forums for supporters to engage, plan rallies and other events and keep all similar posts in one place through Facebook. It also allows administrators of pages to remove and ban dissenting voices, giving a greater degree of control than platforms such as Twitter. In 2016, research on social media usage indicated that 41 per cent of Sri Lankan Facebook users were between the ages of 18 and 24 years (Ishara 2015). Accordingly, a high number of Sri Lankan young social media users risk exposure to hate speech on Facebook (Wickremesinhe and Hattotuwa 2016).

TARGETS AND PRODUCERS

Profiling producers of hate speech on social media is hampered by the lack of verifiable data due to the proliferation of fake accounts and fake identities, particularly on Facebook. Nevertheless, analysis of content, member samples and other data gathered from hate speech disseminating social media sites allows us to make some assumptions.

As noted in previous studies (Samaratunge and Hattotuwa 2014; Wickremesinhe and Hattotuwa 2015, 2016), a very high percentage of hate content on Facebook is posted in the Sinhala language, with comments in Sinhala produced using the English alphabet. This, together with the fact that a high number of users are youth (Ishara 2015) and also samples of accessible data on member pages, such as photographs, interests, euphemisms and language used, indicates that the key followers and producers of hate content are Sinhala-speaking youth. Sentiments expressed through the hate content (both text and graphics) generally indicate radical, ultra-nationalists and sexist ideologies. Among them, there are producers who are firm believers and proponents of such ideologies, as well as followers who simply amplify the majority view in order to 'belong' to the group. There are also followers who are 'converted' to subscribe to the same views and hate the 'other' through constant exposure to the hate content and fear-mongering that usually drives hate speech.

Religious minorities in Sri Lanka, namely Muslims and Christians, were victims of both online hate speech and a high volume of physical violence during 2010–2015 (BBC News 2014; Morning Star News 2014). The nature of online hate content driven by Islamophobia, in particular, during this period has been vicious. It is significant that perpetrators of actual acts of violence against minority Muslim and Christian religious communities during this period went largely unpunished, which emboldened the perpetrators. Online hate speech targeting religious minorities still subsists to varying degrees. Other significant targets include the Tamil ethnic minority, political leaders, including the then President Maithripala Sirisena and the then Prime Minister Ranil Wickremesinghe, women and gay, lesbian and transgender communities. Following the military defeat of the Liberation Tigers of Tamil Eelam by the Sri Lankan military in 2009, which ended 30 years of war and terrorism, a new form of racism came into play, with social media being used to fan the flames of communalism. Reinforcing anti-Tamil sentiments using denialism, 'othering' and scare tactics are common characteristics of hate speech, whereby the Tamil community is labelled as terrorists and any attempt to resolve important issues of contention such as land

ownership is equated with supporting terrorism (Wickremesinhe and Hattotuwa 2016, 15).

Utilising social media as an election campaigning tool is relatively new in Sri Lanka. During the 2015 general election campaign period, Facebook emerged as a widely used space as never before (Gunawardena 2015). Social media analyst Nalaka Gunawardena (2015) notes that the forerunners of social media political discourse 'were not politicians or their support teams but politically charged and digitally empowered citizens, especially youth'. The presidential election of January 2015 and the parliamentary general election of August 2015 witnessed online political engagement as well as an abundance of thinly veiled hate speech, mostly against the contenders: Maithripala Sirisena, presidential candidate of the New Democratic Front led by the United National Party, and Ranil Wickremesinghe, leading the United National Front for Good Governance (UNFGG). For example, in a study of a sample Facebook page conducted in 2015 (Wickremesinhe and Hattotuwa 2015), out of 31 posts by the administrator during a period of one month, 17 posts were directly critical of or inciting hate against the above-mentioned leaders and other politicians of the United National Party, while only 11 posts were non-political and directly dealt with the cause of the page, which was in support of an army soldier who was convicted of the infamous Mirusuvil massacre of civilians and sentenced to death.[1] Of the 17 posts mentioned above, 12 were critical of the government of President Sirisena and his manifesto of *yahapalanaya* (good governance). Five posts were directly critical of Prime Minister Wickremesinghe and the United National Party. While the posts themselves do not always directly amount to hate speech, the comments posted in response do.

It is interesting to note that candidates from minority ethnic and religious groups (Muslim, Christian and Tamil), as well as women candidates, were targeted by hate speech producers during the same period of elections not by reasons of political ideology but simply by their race, religion or gender. A correlation seems to exist between ultra-nationalist ideology and certain political

[1] http://www.dailymirror.lk/77509/8-idps-murdered-in-jaffna-soldier-sentenced-to-death (accessed on 14 May 2016); http://adaderana.lk/news/31381/army-soldier-sentenced-to-death-over-mirusuvil-massacre (accessed on 15 May 2016).

parties and candidates who were in power at the time, whose supporters produced such hate speech. The absence of any retraction or denouncement by them may be construed as their direct or tacit approval, even if they did not personally engage in producing online hate content.

The high volume of hate spewed out by social media and particularly Facebook pages (11 of which were reviewed in a study [Wickremesinhe and Hattotuwa 2016]) did not translate to hate crimes in the form of physical manifestations of violence or reaction against the targets. Writer Nalaka Gunawardene (2015) poses and answers the question:

What role (if any) did social media play in the recently concluded Parliamentary (General) Election on 17 August 2015?

Many are asking this question—and coming up with different answers. That is characteristic of our new reality: there is no single right answer when it comes to multi-faceted and fast-evolving phenomena like social media.

Gunawardene's observation also applies to the role of hate speech on social media; there is no single right answer as to the impact or significance of hate speech online. While the content and volume of online hate speech examined (Wickremesinhe and Hattotuwa 2016) show a clear pattern of targeting candidates from the UNFGG and not the United Peoples Freedom Alliance led by former President Mahinda Rajapaksa, the UNFGG win led by Premier Ranil Wickremesinghe and the comparatively poor performance of the United Peoples Freedom Alliance raises questions if online hate speech had any significant influence on voters.

IS THERE A CAUSAL LINK BETWEEN ONLINE HATE SPEECH AND PHYSICAL VIOLENCE OR REACTION?

The period between 2010 and 2015 witnessed heightened ultra-nationalist activity and violence championed by groups such as the Bodu Bala Sena, which can be translated loosely as 'Buddhist Power Force or Army'. The influence of such groups in physical manifestations of hate targeting ethnic and religious minorities is also mirrored in hate speech targeting on social media. While a direct causal link between online hate speech and physical manifestations of

violence or action propelled by such speech is difficult to establish, there are instances where a somewhat disturbing correlation can be noted between online hate speech and exacerbations of physical violence, as per a study conducted in 2014 (Samaratunge and Hattotuwa 2014, 26). The anti-Muslim riots in Aluthgama are a case in point. In June 2014, Buddhist groups circulated a story online that a monk was attacked by Muslims, sparking a riot in Aluthgama in the south of Sri Lanka where 4 people were killed, 200 homes and shops destroyed and 17 mosques attacked. More than 2,200 people were displaced by the violence (Groundviews 2015). The deluge of online hate speech was supported by extensive coverage of similar rhetoric by fringe groups such as the Bodu Bala Sena, Sinhala Ravaya (loosely translated as the sound of Sinhala) and other groups in mainstream media and at public rallies, inciting violence.

On the other hand, many of the candidates targeted by online hate speech during the 2015 elections were elected to Parliament, including candidates from minority ethnic and religious groups, dismissing any causal link to the outcome of the election. However, the party's political hate speech content seems to be directed at a specific group of voters. According to the Department of Census and Statistics Sri Lanka (2015), Internet usage in Sri Lanka is highest among youth between the ages of 20–24 years (27%) and 25–29-year-olds are the second highest users (25). Registered voters in the 2010 general election totalled 12,071,062, while registered voters in the 2015 general election totalled 15,044,490, indicating a 20 per cent increase, or approximately 2.9 million new voters, over the intervening period of 5 years. This number constitutes youth (between the ages of 18 and 33 years), including first-time voters who were eligible electors at the 2015 general election. It is, therefore, probable that a segment of young voters or first-time voters were exposed to online hate speech. Whether this constitutes a deliberate strategy or an accident is unclear.

THE IMPACT OF SOCIAL MEDIA HATE SPEECH ON SOCIETY: TRENDS AND CHALLENGES

The specific targeting of youth with hate speech—whether by design or accident—is a challenge to any society, as it could potentially

result in serious consequences. Is the proliferation of online hate speech in Sri Lanka a slow-ticking time bomb of youth radicalisation? Radicalisation is the process by which people become convinced that violence against others and even themselves is justified in defence of their own group. Not everyone in a hate group is necessarily radicalised to the same degree, and only a few members may be radicalised to the point where they are ready to advocate and commit violence. In an article titled 'Mechanisms of Political Radicalization',[2] authors Clark McCauley and Sophia Moskalenko identify five ways in which a person or group may be radicalised through online hate speech. As noted in the 'Liking Violence' study (Samaratunge and Hattotuwa 2014, 26) and reiterated by this study, the following are evident in Sri Lanka's online hate groups.

- **Slippery slope:** Journeying from reading hateful content to creating it.
- **The power of love:** The social and emotional bond created by belonging to a group.
- **Radicalisation in like-minded groups:** Where, over time, members are pressured to conform to the radical ideology or opinion of the group, even if they hold different views at the start.
- **Radicalisation under isolation or threat:** The felt need to support a cause or group against a perceived external threat, particularly if one belongs to that group.
- **Dehumanisation:** The portrayal of the opponents as less than human enemies creates a sense of justice in taking any course of action against them.

The danger of youth radicalisation over time is a real concern. The skilful methods utilised by extremist groups for online radicalisation and recruitment of youth have led governments and society at large in many countries to seek creative countermeasures. A Reuter story highlights the out-of-the-box thinking adopted by a French anthropologist who says 'to tackle jihadist, ditch reason' reports

[2] http://bascom.brynmawr.edu/aschcenter/mccauley/webpage%20stuff/2008%20mechanisms%20rad%20McC%20Moskale.pdf (accessed on 09 May 2016).

Pauline Mevel and Chine Labbé.[3] While it is not the intention of this chapter to sound alarmist or propose drastic countermeasures to restrict freedom of online speech, it is prudent to take note of the negative effects of unchecked hate speech on youth and society at large where radicalisation over time may mutate into actual violence, as witnessed in Kenya and Rwanda. Some relevant and thought-provoking statistics taken from the Centre for Policy Alternatives survey on 'Consumption and Perceptions of Mainstream and Social Media in the Western Province' reveals that 61.5 per cent of respondents would share with others content pertaining to political or social issues they read online (Centre for Policy Alternatives 2016). The same study reveals that 37.2 per cent of respondents are more likely to believe online content as truth if they see it first shared by a friend, rather than direct from a media website, and 51 per cent would reconsider their opinion if they see a friend share a story they were previously reluctant to accept as truth. While these statistics are based on a survey of the Western Province, the snapshot of patterns and trends therein provide insights into the usage and influence of social media throughout Sri Lanka.

The nature of the hate speech found on social media groups often mirrors hate speech content posted on other groups that share similar views. Cross-posting is common. Social media such as Facebook groups create virtual communities that 'gather', bond and vent over shared interests or causes. This like-mindedness is the glue that binds group members together. Such groups become echo chambers where an inkling of hate is quickly amplified into something larger by feeding group members information that simply reinforces that initial thought (Koebler 2014). Intolerance of 'outsiders' or those who disagree or challenge the ethos of the group and vicious hate speech targeting them, as exemplified in many social media groups, is common. This creation of a space that reflects one's own ideals with the exclusion of others is similar in effect to a 'filter bubble' (Eli Pariser 2011a), a term coined by the Internet and political activist Eli Pariser to explain the result of a personalised search used by online search engines to guess what information a user would like to see based on information

[3] http://mobile.reuters.com/article/idUSKCN0SS14W20151103?utm_source<hig>=</hig>twitter (accessed on 17 May 2016).

such as past click behaviour and search history. According to Pariser, a user, in effect, becomes less exposed to information or views that contradict their own and isolates users in their own ideological bubbles (Eli Pariser 2011b). Similarly, members of these groups are less likely to be exposed to or accept views contrary to their own.

They are also more likely to accept even the most implausible story as truth and unquestioningly subscribe to extreme views and calls to violence, provided it is shared by what they perceive as a trusted source, a like-minded social media group. Most hate groups utilise graphics to post their messages. This form of non-verbal communication is more efficient and impactful in that it is attractive even to very young viewers and more likely to be viewed, 'Liked', shared and digested by members. Prolonged exposure to violent imagery—photographs, drawings and videos promoting killing, torture, rape or injury targeting a particular community or individual—can desensitise impressionable youth to a point where violence no longer shocks or repels but becomes normal and acceptable. Studies carried out in other countries (the USA and Canada) report children as young as 9 or 10 years old being exposed to online hate and violence (Ybarra et al. 2008). While such a study has not been conducted in Sri Lanka to provide conclusive data, it is safe to venture that in Sri Lanka too, children are exposed to hate content on platforms such as YouTube and Facebook.

CHALLENGES IN COMBATING ONLINE HATE SPEECH

Internet platform providers such as Google, Facebook and Twitter have developed robust guidelines around the content they will allow on their platforms, but these seem to work best around output that is in English. Hate speech existing on social media is often undetected by platform, domain or app owners, such as Facebook's hate-speech monitoring mechanisms[4] due to the expression being predominantly in Sinhala, often using the English alphabet (even the language used to annotate photos, illustrate videos or draw

[4] https://www.facebook.com/notes/facebook-safety/controversial-harmful-and-hateful-speech-on-facebook/574430655911054 (accessed on 18 May 2016).

memes is predominantly, if not exclusively, Sinhala). This clearly lies outside the existing language competencies of Facebook's automated and human-curated monitoring frameworks. This is why content in English runs completely counter to Facebook's policies against hate speech[5] and finds freedom of expression in Sinhala, only subject to scrutiny and compliance when reported by conscientious users. More efficient monitoring and action against online hate speech is needed not only for Sri Lanka, but also for other countries facing similar challenges.

Even where hate speech content is posted in the English language, there are difficulties in effecting guidelines or standards set by platforms, domains or apps in a universal context due to the uniqueness of cultures, whereby what is offensive in one may not be so in another. Content acceptable in or posted lawfully in one country or region can be deemed hateful and unlawful in another, even on the same platform or site.

The most basic definition of hate speech is understood as 'speech that attacks, threatens or insults a person or group on the basis of national origin, ethnicity, colour, religion, gender, gender identity, sexual orientation or disability.' However, definitions of hate speech vary depending on country, region and, in some cases, organisation. The term is, unsurprisingly, variously defined across leading web companies, as evident in their guidelines and boundaries on hate speech, for example, Google's YouTube,[6] Facebook[7] and Twitter.[8] Add to these varying definitions that the challenge of defining hate speech—in contradistinction to, for example, content that is just mildly offensive, distasteful, satirical or acerbic—is deeply rooted in context and expression. What could be a generally accepted turn of phrase used in colloquial speech, when translated into English and out of context, can be seen as hate speech in line with the guidelines noted above. Similarly, deeply offensive or hateful sentiments may be expressed freely, disguised by seemingly innocent colloquial turns of phrase, which when translated into

[5] https://www.facebook.com/communitystandards (accessed on 18 May 2016).

[6] https://www.youtube.com/yt/policyandsafety/communityguidelines.html (accessed on 18 May 2016).

[7] https://www.facebook.com/help/135402139904490/ (accessed on 18 May 2016).

[8] https://support.twitter.com/articles/18311 (accessed on 18 May 2016).

English would seem equally innocent. Numerous such examples where inflammatory and defamatory content against a specific community finds expression and openly resides on Facebook have been documented by the authors.

This brings us to a key challenge in hate speech: It always requires context to understand and address, and increasingly, the intermediaries in both supporting and curtailing its spread are corporate entities, not governments. Machine-level and algorithmic frameworks to identify and block hateful and harmful content often fail simply because they flag too many false positives (content erroneously flagged as hate speech) or allow so much hate speech to pass through (in, as noted earlier, languages other than English) that their core purpose is rendered irrelevant. This puts the burden of addressing this content on the users themselves, who, through reporting mechanisms baked into all the major social media platforms, can choose to report hate speech with relevant context. As these reporting mechanisms are only as effective as the number of people who report hate speech, they also take some time to kick in from the time of submission to the actual deletion or blocking of the original content, page, account or user. There is also no guarantee that the (corporate) owner of an app, service, platform or website agrees with the reporting of hateful content. While Facebook has acted swiftly in some instances to deal with hate speech,[9] studies show significant variance in dealing with hate speech even within Facebook.[10]

Anonymity is one of the most abused privileges online. Hate speech producers online, in particular, are able to hide behind an assumed identity and engage in the most vicious hate speech against communities or individuals, and the victims are left helpless. While some platform providers such as Twitter and Facebook have mechanisms for preventing or removing fake accounts, the Sri Lankan Facebook community has a considerable number of fake account holders. According to the Sri Lanka Police, 20 per cent of Sri Lanka's Facebook accounts, out of a total of 1.2 million (in 2012),

[9] http://www.cnet.com/news/facebook-vows-to-help-germany-combat-racist-xenophobic-content/#ftag<hig>=</hig>CAD590a51e (accessed on 17 September 2015).

[10] http://ohpi.org.au/if-you-cant-recognize-hate-speech-the-sunlight-cant-penetrate/ (accessed on 18 May 2016).

were fake.[11] Sri Lankan authorities recently stepped up measures to deal with increasing numbers of cybercrimes. The Computer Crimes Division reveals that 2,000 complaints involving Facebook and Twitter were reported in the first seven months of 2015.[12] Most of these incidents were related to Facebook, and primarily involved the use of fake profiles.

Another challenge to addressing hate speech is that it is technically impossible—given the volume, variety and velocity of content production on the Internet today—to robustly assess and curtail, in as close to real-time as possible, inflammatory, dangerous or hateful content just in English, let alone other languages like Sinhala or Tamil. Once the content is produced and posted on a single platform, it generates responses and often replicates and mutates into other content, which is again posted on dozens of other websites and platforms. This makes it impossible to completely erase a record of its existence, even if the original is taken down, deleted or redacted. This makes it extremely hard to address the harm arising from hate speech since there is so much of it around in digital form across so many media platforms.

THE IMPACT OF POLITICAL CHANGE

As discussed above, the growth of hate speech online in Sri Lanka poses a range of challenges to government and governance around social, ethnic, cultural and religious co-existence and diversity and, ultimately, to the very core of debates around how we see and organise ourselves post-war. There is simply no panacea, no easy fix or solution in the short term that will effectively curtail the emergence of hate speech online in the future. Indeed, a government that protects instigators of hate is not one that can drive progressive policies to address a growing trend of this same hate expressed online. Politicians who are digitally illiterate are equally ill-placed to bring about legislation that addresses hate speech, even though it may appear to be expedient to do so in light of increasing violence.

---

[11] www.facebook.com/SLtelecom/posts/366614890015926 (accessed on 18 May 2016).

[12] http://www.dailynews.lk/?q<hig>=</hig>2015/11/05/features/new-laws-curb-cyber-crimes-0 (accessed on 18 May 2016).

In 2015, the newly elected Sri Lankan Sirisena–Wickremesinghe government adopted hasty measures to combat hate speech. For the first time, the prime minister spoke strongly against attempts by local print media to fan communal hatred,[13] which in itself can be seen as a positive move, as the government was seen to take a tough stance against hate speech mongering. While there have not been significant strides by this government in combating online hate speech—with the exception of strengthening the Cyber Crime Unit of the Central Investigations Department—the creation of a political and social context which affords less space for impunity is far more likely to significantly reduce the traction of ultra-nationalist groups' fringe lunacy-driven hate rhetoric in both real space and cyberspace.

> Certain utterances achieve terrifying power, in the right context. In a climate of ethnic animosity, statements of ethnic pride are indistinguishable from insults against one's opponents. And the converse is also true: Even the most hateful or inciteful speech remains benign, if it has no audience or if its audience is firmly and explicitly determined to keep the peace (Hirsch, quoted in USHMM, n.d.)

THE LAW VERSUS FREEDOM OF EXPRESSION

Ultimately, there is no technical solution to what is a socio-political problem. Sri Lanka's culture of impunity and the breakdown in the rule of law that prevailed during the 2010–2015 period are what afford the space for radical groups such as the Bodu Bala Sena, Sinhala Ravaya and Ravana Balakaya to engage in verbal and physical campaigns of racial and religious intolerance targeting minorities. The hate that was produced on social media mirrored and amplified the unchecked hate that was visible in real life, which seemed to have the tacit approval of authorities. The Sri Lankan legal system affords a range of legal remedies and frameworks to hold perpetrators of hate speech accountable for their violence, whether verbal or physical. The issue is not the non-existence of relevant legal frameworks, but their non-application or selective application.

---

[13] https://archive2.srilankamirror.com/news/item/2340-ranil-gets-tough-with-the-media (accessed on 18 May 2016).

The debate as to whether Sri Lanka needs hate speech laws gathered momentum in 2015, resulting in a hasty decision by the government of Sri Lanka to criminalise hate speech by amending the penal code. However, the proposed amendments[14] were shelved in December 2015,[15] bowing to pressure from legal and human rights activists that the proposed legislation posed a real danger of being misinterpreted and misused as a tool of harassment and repression. It is a very thin line that divides political speech from hate speech. The introduction of new legal measures needs careful consideration. Further, as indicated above, the existing legal framework, if applied without bias, is sufficient in dealing with hate speech. For example, in late April 2016, a man who posted a video on Facebook allegedly calling on the security forces and the police to execute the president and the prime minister was arrested and remanded (Thaudeen 2016).

The then United Nations special rapporteur for the promotion and protection of the right to freedom of opinion and expression, Frank La Rue, in his report to the United Nations General Assembly Human Rights Council (2011), stated that 'while states are the duty-bearers for human rights, private actors and business enterprises also have a responsibility to respect human rights'. In this regard, the special rapporteur highlights the framework of 'protect, respect and remedy'.

> The framework rests on three pillars: (a) the duty of the State to protect against human rights abuses by third parties, including business enterprises, through appropriate policies, regulation and adjudication; (b) the corporate responsibility to respect human rights, which means that business enterprises should act with due diligence to avoid infringing the rights of others and to address adverse impacts with which they are involved; and (c) the need for greater access by victims to effective remedy, both judicial and non-judicial.

[14] The amendment sought to introduce Section 291C to the Penal Code, No. 11 of 1887. The proposed Section 291C was nearly identical to Section 2(1)(h) of the controversial Prevention of Terrorism (Temporary Provisions) Act, No. 48 of 1979 (PTA).

[15] https://asianmirror.lk/index.php?option=com_k2&view=item&id=13499:government-backtracks-on-amendment-to-penal-code&Itemid=114 (accessed on 18 May 2016).

To counter hate speech and protect the vulnerable, the multidimensional and complex nature of hate speech requires collaboration among government, platform providers and the empowerment of users. Countermeasures, education and empowerment of users by the state and civil society, as well as more diligent monitoring by platform providers, are some of the useful measures that can be adopted in combating online hate speech.

Although there is no easy, single or prescribed solution, progressive thinking, proactive content production, strategic interventions and careful monitoring can identify and neutralise the wider harm online hate speech can cause if allowed to grow unchecked. A sound legal framework applied by independent law enforcement and the judiciary can deal with perpetrators who promote, produce and disseminate hate speech. A system where legal and non-legal measures work together balance the right to freedom of expression and eradicate the scourge of online hate speech.

If the trend of hate speech online is to be truly stemmed, media literacy programmes aimed at students, teachers, parents, lecturers, government officials, civil society, media consumers and citizens, in Sinhala and Tamil, in mobile-friendly, freely accessible and engaging ways are required over the long-term. Civil society must engage in producing and promoting counter-messaging to strategically counter hate speech on social media. This will also require civic media initiatives and citizen journalists to tie up with principled professional journalists working in the mainstream media to fact-check, debunk, engage and dispel rumours on social media in close to real-time—vital skills that are aided by constantly improving digital media literacy.

RECOMMENDATIONS

Some recommendations, for Sri Lanka and beyond, around countering violent extremism and hate speech online are as follows:

1. **Counter-messaging to stymie spread of hate:** Counter-messaging is the production of content around common vulnerabilities and exposures (CVE) and disseminated using the same platforms, apps, services and sites as more harmful content. This can include, on a case-by-case basis, direct engagement

with accounts that promote violent extremism by debunking misinformation and disinformation campaigns and calling their bluff on pseudo-science and myth-making.

2. **Upscaling civil society capacity:** As noted by Hattotuwa in 2013,[16] 'groups which attempt to portray a more inclusive and tolerant country, by critiquing the positions of the extremists, often come under attack, are subject to hate speech and fail to attract as many followers as the Facebook pages and groups with inflammatory content'. This in turn calls for sufficient human, financial and institutional resources to support more sustained monitoring of these hate speech trends, so as to create early warning mechanisms that alert relevant authorities and civil society stakeholders about heightened tensions online that could lead to, or exacerbate, real-world violence. Importantly, this monitoring should cover vernacular languages in addition to English content across key social media domains, apps, services and platforms.
3. **Ensuring measures to address violent extremism online don't impinge on rights:** Along with more sustained, deeper monitoring of violent extremism online and digital media literacy campaigns around the critical appreciation of such content geared towards those between 18 and 30 in particular, recommendations to address hate speech online echo points made in the BytesForAll report on Pakistan's hate speech in cyberspace,[17] pegged to

> A multi-pronged approach to work, a plan of action that has multiple stakeholders involved would be necessary to maintain checks and balances, particularly to ensure that the issue of hate speech in cyberspace is not manipulated and used to further political agendas, increase censorship and/or target and discriminate against vulnerable individuals/groups.

BytesForAll focuses on the role of the government, mainstream media, online companies (e.g., Facebook) and organised civil society advocacy and activism as means through which online

[16] https://sanjanah.wordpress.com/2013/02/01/anti-muslim-hate-online-in-post-war-sri-lanka/ (accessed on 19 May 2016).

[17] https://www.scribd.com/document/257947349/Pakistan-Hate-Speech-Report-2014 (accessed on 19 May 2016).

hate speech can be, to the extent possible, effectively contained and addressed.

4. **Voluntary codes of conduct:** However, especially in contexts where there is a democratic deficit, the challenge of engaging with the government and informing progressive policymaking is even more acute, if not downright impossible. The mainstream media, in general, is extremely risk averse and has neither the imagination nor independence to counter hate speech by extremist groups, especially when it is widely perceived to be protected by powerful sections of the government. Furthermore, little to no comment moderation guidelines across mainstream media websites also result in trolls openly publishing comments full of hate, hurt and harm. In this light, comment moderation and content curation policies in line with what the World Editors Forum published in 2013[18] can greatly contribute to the creation of official mainstream media websites and social media accounts that actively resist and combat hate speech and engage in CVE, complementing editorial policies that also, on principle, disallow defamatory and inflammatory content from the institution's articles, columns and broadcasts.
5. **Public pledges against hate speech:** There are also other possibilities arising from Dr Tarlach McGonagle's work on addressing online hate speech in Europe.[19] One of his most important ideas, and one that is well worth embracing, is to develop and effectively promote an 'Anti-hate Speech Pledge' for politicians and political parties. As noted by Dr McGonagle,

> A certain minimum number of commitments (around combatting hate speech) would have to be entered into, in return for which, a party could display the logo for the Pledge on all of its official materials. In order to ensure seriousness of purpose and meaningful uptake, participating party leaders would be obliged to attend annual meetings to explain and evaluate their parties' actions to combat hate speech. A non-roll-back clause could be included in order to ensure that annual achievements would continuously be built on.

[18] https://netino.fr/wp-content/uploads/2013/10/WAN-IFRA_Online_Commenting.pdf (accessed on 19 May 2016).

[19] https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId<hig>=</hig>09000016800c170f (accessed on 19 May 2016).

It is clearly in the interests of all political parties to explore ways in which their party leadership, officer bearers and supporters can counter hate speech, in general, and online hate speech, in particular, by signing a pledge that includes public and visible punitive measures against anyone who goes on to produce or disseminate hate and harm. A political culture of zero tolerance for hate speech can deeply influence the production and appraisal of hate speech online.

6. **Demographics are important:** Youth (aged 18–24) stand the risk of radicalisation upon entering and engaging with online and mobile chat-based fora. To appeal to this segment, iconic figures from youth (singers, actors, sportspersons, YouTube producers, hackers, information and technology industry leaders and young entrepreneurs) are more important to leverage in counter-speech initiatives than expressions from or iconography based around *dhamma* (Buddhist scripture). It is also the case that, combined with geo-targeting, those who are held in high regard by this age group in local communities (ranging from monks in a community temple where this segment has gone for tuition or Sunday school to local business owners) can be leveraged, the emphasis being on the identification of influencers within that demographic and, furthermore, by geography.
7. **Geo-targeting/geo-fencing:** Easily done on Facebook, counter-speech content (ranging from pages to specific posts on Facebook) can be targeted to specific regions, at specific times, for specific communities. Wide-scatter promotions simply don't work, either displaying on the screens of those who are already partial to the counter-speech content, or only sporadically appearing on the screens of those for whom it is most relevant. The larger the terrain of an audience, the greater the emphasis should be on geo-fencing counter-speech content. For example, during an election, constituencies that have witnessed heightened communal or partisan violence can be targeted well before the day of the election with counter-speech messaging to prevent the spread of rumours and other inflammatory content.
8. **Language:** In a multilingual country, CVE is largely ineffective if it isn't conducted in the language that dangerous and hate-speech forums use in their interactions. Hate and dangerous

speech on Facebook often exists only in the vernacular, and counter-speech initiatives in English alone have no relevance or traction. Iconic counter-speech examples like Panzagar in Myanmar can be very effective since they transcend the barriers of language. Short-form video can also be a powerful vector for counter-speech to reach target audiences without necessarily being anchored to a single language.

9. **Translation:** Good translations of CVE content that communicate ideas and meaning are hard to come by, and good translators are generally overworked. Idioms, nuances, aphorisms and adages in languages differ, and native speakers of the language counter-speech content was originally produced in or for and conversant in the language into which it will be translated are hard to come by.
10. **Time:** CVE is a long-term process, and timing is important as far as what is expected as a result. Counter-speech to address and reduce electoral violence requires a different timeline than content that seeks to address deep-rooted communal or religious tension. Project-oriented counter-speech campaigns, which are often driven by relatively short-term funding opportunities, are often too short for any meaningful impact.
11. **Reasons for (social media) engagement:** CVE proponents need to do far more and better research around why, and at what times, hate and dangerous speech content is produced and received with high levels of engagement. What drives the production cycles? Are there links to key political or cultural events? Is there a connection between the utterances of key individuals and the production of hate speech in online forums? Is there a connection between the speeches of political groups, politicians, religious leaders or other individuals and the engagement of online users using dangerous speech? Does hate speech increase in the lead-up to an election, and if so, at what key points?
12. **Law of diminishing returns:** CVE proponents need to create content that doesn't just go viral once. They should also keep in mind that content addressed to the same demographic will, unless very inventive, generate progressively less interest and interaction over time. The higher frequency of content

production sometimes risks the perception of counter-speech as spam, whereas infrequent production also risks ineffective audience engagement. Context is critical to content.

CONCLUSION

This chapter is a brief survey of the more complex and, indeed, changing landscape of online discourse. Shaping the currents of these conversations over text, photography, memes, video, animation, graphics, emoticons and other means on a diverse range of apps, platforms and services are the older fault lines of caste, class, religion, economics and politics. Online, identity is both fluid and entrenched. With anonymity comes the courage to bear witness to inconvenient truths, as well as the ability to spread hate and incite violence. Whether used to denounce, decry and destroy, or to push forward the better angels of our nature, digital discourse will always reflect and contest the meta-narratives that govern our social and political engagements, notions of identity, belonging and, by extension, the very fibre of our democratic institutions and culture.

Tellingly, civility, tolerance and respect for diversity are as hard to find online as they are in Sri Lanka's mainstream party political framework, even post-war. It would be a tragedy if the country's only remaining spaces to ideate, critically reflect and robustly debate, which are online, are taken over by hate-mongers to the extent they are allowed to do so in the real world. So many in Sri Lanka, in various ways, resist violence, whether verbal or physical. The challenge is to strengthen their voices and efforts in light of what is becoming a growing trend of hate speech production online, which, though by no means easy, is also not an insurmountable one.

REFERENCES

BBC News. 2014. 'Sri Lankan Muslims Killed in Aluthgama Clashes with Buddhists. Available at: http://www.bbc.com/news/world-asia-27864716 (accessed on 14 May 2016).

Centre for Policy Alternatives. 2016. *Consumption and Perceptions of Mainstream and Social Media in the Western Province*. Available at: http://www.cpalanka.org/full-report-consumption-and-perceptions-of-mainstream-and-social-media-in-the-western-province/ (accessed on 20 May 2016).

Department of Census and Statistics, Sri Lanka. 2015. *Computer Literacy Statistics—2015 (First Six Months)*. Available at: http://www.statistics.gov.lk/Resource/en/ComputerLiteracy/Bulletins/2015-FirstSixMonths.pdf (accessed on 08 February 2016).

Eli Pariser. 2011a. *The Filter Bubble: What the Internet Is Hiding from You*. New York: Penguin Group.

———. 2011b. 'Beware Online "Filter Bubbles".' Ted Talk. Available at: http://www.ted.com/talks/eli_pariser_beware_online_filter_bubbles?language<hig>=</hig>en (accessed on 17 May 2016).

Gagliardone, Iginio, Danit Gal, Thiago Alves, and Gabriela Martinez. 2015. *Countering Online Hate Speech: UNESCO Series on Online Freedom*, 13. Paris: UNESCO.

Groundviews. 2015. 'Online Hate Speech Stokes Fear of Religious Violence in Sri Lanka.' Available at: http://groundviews.org/2014/09/26/online-hate-speech-stokes-fear-of-religious-violence-in-sri-lanka/ (accessed on 11 May 2016).

Gunawardena, Nalaka. 2015. 'Sri Lanka Parliamentary Election 2015: How Did Social Media Make a Difference? Available at: http://groundviews.org/2015/09/03/sri-lanka-parliamentary-election-2015-how-did-social-media-make-a-difference/ (accessed on 14 May 2016).

Hirsch, Susan quoted in United States Holocaust Memorial Museum (USHMM). n.d. Hate Speech and Group Targeted Violence: The Role of Speech in Violent Conflict. Available at: http://www.genocide-watch.com/images/OutsideResearch_Hate_Speech_and_Group-Targeted_Violence.pdf (accessed on 20 February 2016).

Ishara, S. 2015. 'Five Unconventional Social Media Trends in Sri Lanka.' Available at: https://roar.media/english/tech/insights/five-unconventional-social-media-trends-in-sri-lanka (accessed on 08 February 2016).

Keizer, G. 2014. *Getting Schooled: The Reeducation of an American Teacher*. New York: Metropolitan Books. Available at: https://www.goodreads.com/work/quotes/26123571-getting-schooled-the-reeducation-of-an-american-teacher (accessed on 12 May 2016).

Koebler, Jason. 2014. 'The State Department Is Worried Social Media Bubbles Escalate Social Unrest.' Available at: http://motherboard.vice.com/read/the-state-department-is-worried-social-media-bubbles-escalate-social-unrest?trk_source<hig>=</hig>recommended (accessed on 17 May 2016).

Morning Star News. 2014. 'Buddhist Nationalism Continues to Haunt Christians in Sri Lanka.' Available at: http://morningstarnews.org/2014/03/buddhist-nationalism-fuels-attacks-on-christians-in-sri-lanka-2/ (accessed on 14 May 2016).

Samaratunge, Shilpa, and Sanjana Hattotuwa. 2014. 'Liking Violence.' Centre for Policy Alternatives. Available at: https://www.cpalanka.org/wp-content/uploads/2014/09/Hate-Speech-Executive-Summary.pdf (accessed on 26 April 2022).

Thaudeen, T. Farook. 2016. 'Man Who Allegedly Threatened President, PM on FB Remanded.' Daily Mirror, 29 April, A6.

United Nations General Assembly Human Rights Council. 2011. *Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Frank La Rue*. Article 45, 13. Available at: http://www2.ohchr.org/english/bodies/hrcouncil/docs/17session/A.HRC.17.27_en.pdf (accessed on 18 May 2016).

Wickremesinhe, Roshini, and Sanjana Hattotuwa. 2016. 'Voting in Hate: A Study of Hate Speech on Facebook Surrounding Sri Lanka's Parliamentary Election of 2015.' Centre for Policy Alternatives. Available at: http://www.cpalanka.org/voting-in-hate-a-study-of-hate-speech-on-facebook-surrounding-sri-lankas-parliamentary-election-of-2015/ (accessed on 14 May 2016).

———. 2015. 'Saving Sunil.' Centre for Policy Alternatives. Available at: https://www.cpalanka.org/wp-content/uploads/2015/10/SS-Final-RW-SH-formatted.pdf (accessed on 26 April 2022).

Ybarra, Michele L., Marie Diener-West, Dana Markow, Philip J. Leaf, Merle Hamburger, and Paul Boxer. 2008. 'Linkages between Internet and Other Media Violence with Seriously Violent Behavior by Youth.' *Pediatrics* 122, no. 5 (November). MediaSmarts. n. d. 'For Parents.' Available at: http://www.bewebaware.ca/english/violent_hateful_content.html (accessed on 18 May 2016).