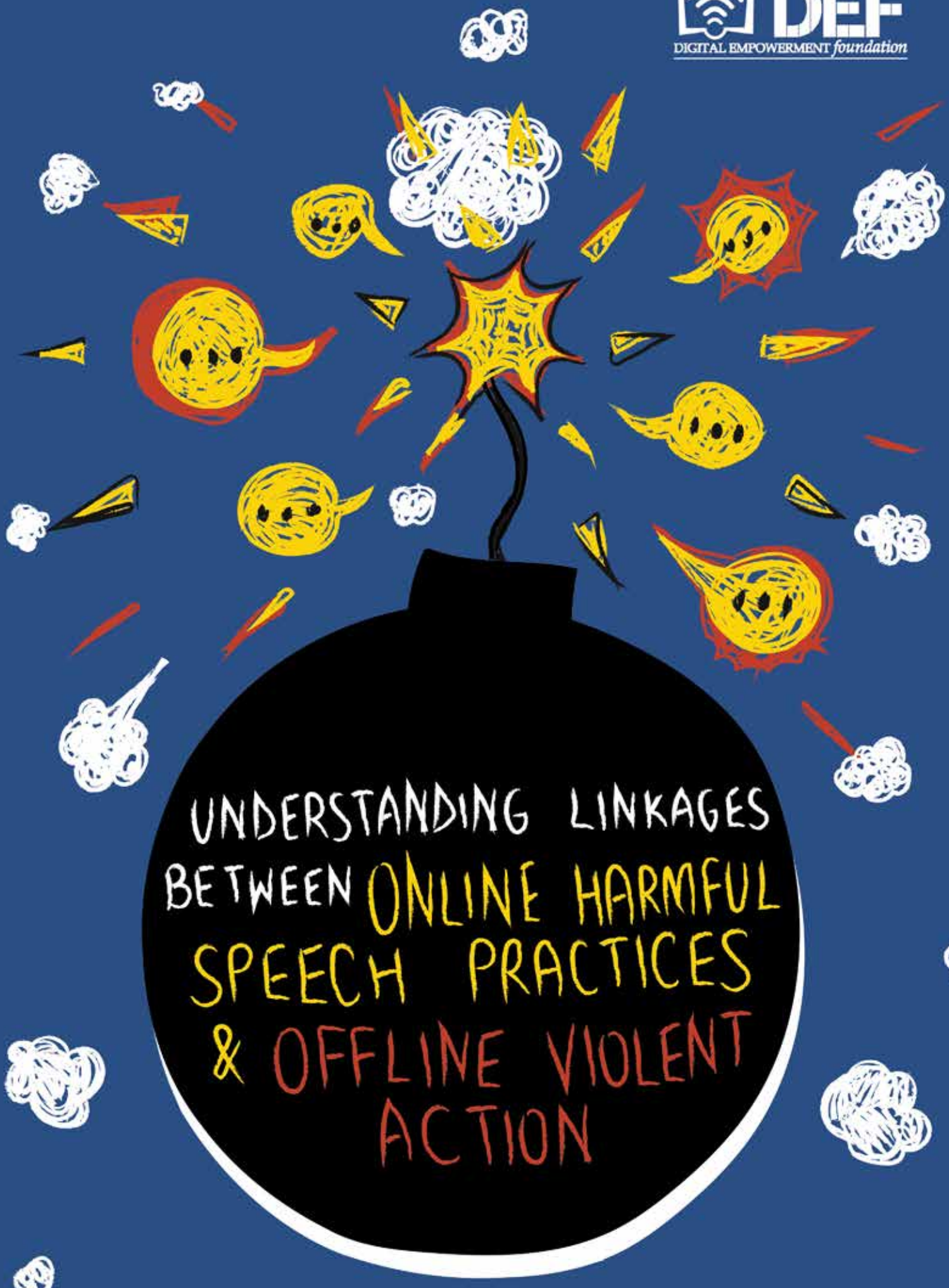
DEF
DIGITAL EMPOWERMENT *foundation*



UNDERSTANDING LINKAGES BETWEEN ONLINE HARMFUL SPEECH PRACTICES & OFFLINE VIOLENT ACTION

Understanding Linkages Between Online Harmful Speech Practices And Offline Violent Action

2020

This work is licensed under a creative commons Attribution 4.0 International License.

You can modify and build upon this document non-commercially, as long as you give credit to the original authors and license your new creation under the identical terms.

Suggested citation:

Content Policy Research Group. (2020). Understanding linkages between online harmful speech practices and offline violent action. Digital Empowerment Foundation: New Delhi.

You can read the online copy under Reports Section at https://www.defindia.org/publications/

Digital Empowerment Foundation
House no. 44, 2nd and 3rd Floor (next to Naraina IIT Academy)
Kalu Sarai (near IIT Flyover)
New Delhi – 110016
Tel: 91-11-42233100 / Fax: 91-11-26532787
Email: def@defindia.net | URL: www.defindia.org

CONTENTS

FUNDING DISCLAIMER

In 2019, the Digital Empowerment Foundation was one of the recipients of Facebook Content Policy Research Awards to understand the linkages between hate speech and offline violence in India.

SITUATING THE RESEARCH – THE INDIAN CONTEXT

In 2020 over 50% of the 1.3 billion Indian population has access to social media; this is up from 19% in 2015, 30% in 2017, and 46% in 2019[1]. This rapid social media penetration is said to be the product of cheap mobile data and increased smartphone penetration[2]. The average Indian social media user spends more than 17 hours on the platforms each week which is more than social media users in China and United States[3].

However, along with this digital transformation – over the past few years, India has seen the alarming rise in violence that have been linked to social media. Affordances provided by social media in the form of distributed democratic and civic participation has also provided the fertile ground for online vitriol, social panic, and targeted campaigns of intimidation and harassment.

The years 2017 and 2018 saw an alarming rise in mob violence pan-India fueled by rumours circulated via social media[4] related to child-lifting, organ-harvesting, and cow slaughter[5]. With regard to child – lifting rumours, despite the repeating trends across the vast expanse of the country they were primarily driven around possible presence of outsiders in the community who were touted as child lifters. This is despite there being no instances of kidnappings being reported or suspected in the lynching spots in the previous 3 months[6].

In tandem with the mob violence around child-lifting rumours, there was a similar peak of violence associated with cow-protection vigilantism with 2017 being the worst year with highest number of casualties[7]. The April 2020 lynching of Hindu priests in Palghar, Maharashtra and Bulandshahr, Uttar Pradesh on rumours of thieving provided a grim reminder of India's recent history with misinformation and offline violence[8]. The unfolding of the COVID – 19 pandemic and its impact on social and economic life has further served to calcify identity-based divisions in the country[9].

The implication of the private messaging service towards the perpetration of localized and even large scale civic violence led to government cognizance of its role. In the wake of the incidents of lynching, the Ministry of Electronics and Information Technology (MEITY) wrote to WhatsApp to "take immediate steps to tackle the menace of misuse of their platform wherein inflammatory messages were circulated that led to unfortunate incidents"[10]. In subsequent notices, the government stated the requirements to institute traceability on the platform while have been the subject of multiple petitions concerning related primary and ancillary issues[11].

[1]Statista. (July 2020). India social network penetration 2015-2025. Retrieved from https://www.statista.com/statistics/240960/share-of-indian-population-using-social-networks/ [25 October 2020].

[2]Venkatramakrishnan, R. (30 September 2018). India wants WhatsApp to break encryption and trace inflammatory messages. Should it?. Scroll.in. Retrieved from https://scroll.in/article/895645/india-wants-whatsapp-to-break-encryption-and-trace-inflammatory-messages-should-it [25 October 2020].

[3]McKinsey Global Institute. (2019). Digital India: Technology to transform a connected nation. Retrieved from https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Digital%20India%20Technology%20to%20transform%20a%20connected%20nation/MGI-Digital-India-Report-April-2019.pdf [25 October 2020].

[4]See Saldanha, A., Rajput, P., & Hazare, J. (2018). Child-Lifting Rumours: 33 Killed In 69 Mob Attacks Since Jan 2017. Before That Only 1 Attack In 2012. IndiaSpend. Retrieved from https://www.indiaspend.com/child-lifting-rumours-33-killed-in-69-mob-attacks-since-jan-2017-before-that-only-1-attack-in-2012-2012/ [4 May 2020]

[5]Madrigal, A.C. (25 September 2018). India's lynching epidemic and the problem with blaming tech. The Atlantic. Retrieved from https://www.theatlantic.com/technology/archive/2018/09/whatsapp/571276/ [25 October 2020].

[6]IndianExpress. (15 July 2018). Murderous mob — 9 states, 27 killings, one year: And a pattern to the lynchings. Retrieved from https://indianexpress.com/article/india/murderous-mob-lynching-incidents-in-india-dhule-whatsapp-rumour-5247741/ [4 May 2020].

[7]Saldanha, A. (2017). 2017 Deadliest Year For Cow-Related Hate Crime Since 2010, 86% Of Those Killed Muslim. IndiaSpend. https://archive.indiaspend.com/cover-story/2017-deadliest-year-for-cow-related-hate-crime-since-2010-86-of-those-killed-muslim-12662. [4 May 2020]

[8]PTI. (2020). Three More Cops Suspended in Connection With Mob Lynching of Sadhus in Palghar. News18. Retrieved from https://www.news18.com/news/india/three-more-cops-suspended-in-connection-with-mob-lynching-of-sadhus-in-palghar-2598035.html [4 May 2020]; Sharma, S. (2020). 2 sadhus killed inside Bulandshahr temple in UP, accused arrested. IndiaToday. Retrieved from https://www.indiatoday.in/crime/story/2-sadhus-killed-inside-bulandshahr-temple-in-up-accused-arrested-1671948-2020-04-28 [4 May 2020].

[9]Nizamuddin, F. (2020). The COVID-19 pandemic and the infrastructure of hate in India. Rosa Luxemburg Stiftung. Retrieved from http://www.irgac.org/2020/07/22/the-covid-19-pandemic-and-the-infrastructure-of-hate-in-india/ [25 October 2020].

[10]PIB. (20 July 2018). WhatsApp told to find more effective solutions. Retrieved from https://pib.gov.in/PressReleaseIframePage.aspx?PRID=1539410 [25 October 2020].

[11]Agarwal, A. (31 January 2020). Supreme Court directs Madras HC to transfer all files in the WhatsApp traceability case. Medianama. Retrieved from https://www.medianama.com/2020/01/223-supreme-court-to-madras-hc-transfer-all-files-in-whatsapp-traceability-case/ [25 October 2020].

In response to the rising concerns, WhatsApp limited the number of forwards and appointed a grievance officer for India after the government and the Supreme Court demanded that it improve its approach to safety[12]. In a wider regulatory move, MEITY released the draft Intermediary Guidelines (Amendment) Rules, 2018 under Section 79 of the Information Technology Act, 2000. The Rules were targeted at expanding the liability of social media intermediaries for content hosted on their platforms[13].

The 2018 Rules mandated a 72-hour window of traceability on receiving complaint from law enforcement; disable access within 24-hours to content deemed defamatory, against national security, or in violation of Article 19(2) of the Indian Constitution; requirement for all platforms with more than 5 million users to have a registered office in India[14]. It also included the deployment of automated tools to detect and take down 'unlawful' content[15]. In January 2020, it was reported that the government plans to change the proposed rules such that monitoring and take down measures apply only to big social media companies[16].

Despite the centrality of WhatsApp within India's content policy debate and its implication within adverse social phenomenon – more recently Facebook's murky history within the Myanmar and Sri Lanka India violence echoed within the context of the North – East Delhi riots. India is the biggest market for Facebook - it has 260 million active users in India which is the highest they have in any country in the world with 52% of Indians using Facebook as a source of news[17].

The Delhi Assembly Committee on Peace and Harmony issued summons to Facebook over its alleged "deliberate and intentional" inaction to contain hateful content during the riots that ravaged North-East Delhi in February 2020 and left over 50 people dead[18]. On 31 August 2020, the Committee said that it seemed prima facie that Facebook has a role in the riots and should be treated as a co-accused[19].

Facebook India's Vice-President and Managing Director Ajit Mohan refused to appear before the Delhi Assembly Committee citing two primary reasons: (i) regulation of intermediaries like Facebook falls within the purview of the Central Government; (ii) the subject of law and order in Delhi also falls within the purview of the Central Government[20]. Facebook further filed a plea before the Supreme Court in connection to the summons which then asked the Delhi committee to halt any coercive action against Mohan till 15 October 2020[21].

[12]Venkatramakrishnan, R. (30 September 2018). India wants WhatsApp to break encryption and trace inflammatory messages. Should it?. Scroll.in. Retrieved from https://scroll.in/article/895645/india-wants-whatsapp-to-break-encryption-and-trace-inflammatory-messages-should-it [25 October 2020].

[13]ETech. (09 January 2020). MeitY: Big social media firms to face tougher online content regulation norms. ET Government.com. Retrieved from https://government.economictimes.indiatimes.com/news/digital-india/meity-big-social-media-firms-to-face-tougher-online-content-regulation-norms/73167509 [25 October 2020].

[14]Ibid.

[15]Digital Empowerment Foundation. (2019). Submission of comments on MEITY's draft Information Technology [Intermediary Guidelines (Amendment) Rules], 2018. Retrieved from https://www.defindia.org/wp-content/uploads/2019/04/2.-MEITYs-IT-rules.pdf [25 October 2020].

[16]ETech. (09 January 2020). MeitY: Big social media firms to face tougher online content regulation norms. ET Government.com. Retrieved from https://government.economictimes.indiatimes.com/news/digital-india/meity-big-social-media-firms-to-face-tougher-online-content-regulation-norms/73167509 [25 October 2020].

[17]Sannam S4. (2020). Top social media trends in India in 2020. Retrieved from https://sannams4.com/top-social-media-trends-in-india-2020/ [25 October 2020].

[18]Staff Reporter. (15 September 2020). Attempt to hide crucial facts on Facebook's role in Delhi riots, says Assembly committee. The Hindu. Retrieved from https://www.thehindu.com/news/cities/Delhi/attempt-to-hide-crucial-facts-on-facebooks-role-in-delhi-riots-says-assembly-committee/article32608982.ece [25 October 2020].

[19]Ibid

[20]Jain, P. (15 September 2020). Delhi riots: Facebook skips assembly meet, says city's law and order within Centre's domain. IndiaToday. Retrieved from https://www.indiatoday.in/india/story/facebook-skips-delhi-panel-meet-delhi-law-and-order-within-centre-domain-raghav-chadha-1722000-2020-09-15 [25 October 2020].

[21]Mathur, A. (23 September). Delhi riots: Facebook India gets breather as SC asks Assembly panel to halt coercive action. IndiaToday. Retrieved from https://www.indiatoday.in/india/story/delhi-riots-facebook-india-vc-supreme-court-delhi-assembly-panel-summon-1724609-2020-09-23 [25 October 2020].

REGULATION OF HATE SPEECH AND ITS DISCONTENTS

Hate speech law has been mired within positions held by opponents and proponents of hate speech regulation. Proponents believe such regulations are indispensable for all sections of the population to enjoy their constitutionally guaranteed equality and freedom. Opponents argue that such regulations serve counter to individual liberty, autonomy, free and unfettered participation in democratic life and prevent the formation of public opinion. Proponents on the other hand believe that regulation of hate speech is required to safeguard substantive autonomy, ensure freedom from oppression, guarantee public assurance of civic dignity, ensure recognition of cultural identity, and facilitate real access to participation in democratic life[22].

Despite countries having provisions on hate speech, these discontents play out in their enforcement. This is particularly true for social media platforms hosting user-generated content through self-regulatory forms of self-governance like community guidelines. These regulate the use of such platforms and the speech acts on such platforms. However, despite these irreconcilable differences studies of hate speech has often been divorced from its social implications[23] which has confounded the parameters of its enforcement and entrenched the long-standing social dilemmas on the knock-on effects of regulating hate speech on an individual freedom of expression.

However, establishing a direct causal relationship between hate speech and violence is fraught with complications. Since they are inextricably entwined with structures of power within social relations distributed socially rather than individually. This is particularly why constitutional frameworks based on liberal individualism find it difficult to identify the loci of harm and its perpetration with regard to hate speech. This problematic is further deepened due to lack of deeper social investigation, beyond legal-philosophical frameworks, into the social praxis of hate speech prevalent in society and the violence experienced by it.

In the Indian legal corpus while explicit mention of hate speech is rare, the rationale for its regulation is rooted within the colonial articulations of containing uprisings against the colonial state[24]. Bhatia argues how substantive provisions of the Indian legal regime pertaining to the regulation of hate speech has often used to foreclose the space of civic participation[25]. With regard to sections like 295A (insulting religions or religious feelings), 153A (promoting 'disharmony', 'enmity', 'ill-will', or 'hatred' between different religious groups, castes, communities etc.), 298 (uttering words with deliberate intent to wound the religious feelings of a person); 509 (word, gesture, or act intended to insult the modesty of a woman); 508 (act caused by inducing a person to believe that he will be rendered an object of divine displeasure); 504 (intentional insult with intent to provoke a breach of peace) of the Indian Penal Code, 66A of the Information Technology Act (online censorship)[26] he mentions:

> However, establishing a direct causal relationship between hate speech and violence is fraught with complications. Since they are inextricably entwined with structures of power within social relations distributed socially rather than individually. This is particularly why constitutional frameworks based on liberal individualism find it difficult to identify the loci of harm and its perpetration with regard to hate speech.

[22]Brown, A. (2015). Hate speech law: A philosophical examination. Routledge: New York and London.

[23]Wilson, A. (2019). The digital ethnography of law: Studying online hate speech online and offline. Journal of Legal Anthropology, 3(1), 1-20.

[24]Narrain, S. (2016). Harm in hate speech laws: Examining the origins of the hate speech legislation in India. In S. D. R. Ramdev (Ed.), State of hurt: Sentiment, politics, censorship, pp. 39–54. SAGE Publications: New Delhi, India.

[25]Bhatia, G. (2016). Offend, shock, or disturb: Free speech under the Indian Constitution. Oxford University Press: New Delhi.

[26]Section 66A of the IT Act, 2000 was struck down as unconstitutional in the case of Shreya Singhal v. Union of India on grounds of violating freedom of speech guaranteed under Article 19(1)(a) of the Indian Constitution.

Under these sections, books have been banned; books have been withdrawn; people arrested for political satire, for political critique, and for 'liking' someone else's political critique on Facebook. Requirements of prior sanction and other safeguards ensure that not all cases come to trial. Nonetheless, a large part of the problem is the cognizable nature of these offences under the Code of Criminal Procedure, which grants the police the powers of arrest without the need for obtaining a judicially sanctioned warrant.

From the above passage it would seem that the collection of hate speech regulations have served to undermine the inclusive space envisioned by its proponents. Thereby, serving to reproduce the conditions of power that hate speech regulation hopes to hold in check.

Brown disaggregates hate speech law into 10 clusters: (i) group defamation; (ii) negative stereotyping or stigmatization; (iii) expression of hatred; (iv) incitement to hatred; (v) threats to public order; (vi) denying, etc. acts of mass cruelty, violence, or genocide; (vii) dignitary crimes or torts (like the use of racist language or slurs that are used the undermine the dignity of groups or classes of persons with ascriptive characteristics like race, ethnicity, religion, nationality etc.); (viii) violation of civil and human rights (like the 'right to non-discrimination, the right to fair accommodation, and the right not to be exposed to discriminatory harassment'); (ix) expression oriented hate crimes (like the cross-burning practice of Klu Klux Klan or the use of Nazi symbolism); (x) time, place, and manner restrictions (establishing norms of appropriate action – e.g. constraining protests at given times and locations)[27].

Hate speech discourse pre-determines the effects of hate speech as negative and damaging leading to the regulatory rationale of control and containment. This regulatory effort includes both the state through notified laws and regulations as well as the social media intermediaries through their self-regulatory codes like community guidelines[28]. These clusters reflect practices of exclusion and marginalization that hate speech regulation seeks to negotiate. Social research on hate speech has attempted to understand their prevalence and crystallize them to frameworks of understanding that can aid analysis and observation of such within social reality.

[27]Brown, A. (2015). Hate speech law: A philosophical examination. Routledge: New York and London.
Pohjonen, M. & Udupa, S. (2017). Extreme speech online: An anthropological critique of hate speech debates. International Journal of Communication, 11(2017), 1173 – 1191.

[28]Pohjonen, M. & Udupa, S. (2017). Extreme speech online: An anthropological critique of hate speech debates. International Journal of Communication, 11(2017), 1173 – 1191.

UNDERSTANDING THE ONLINE PRACTICE OF HATE SPEECH

Definitional challenges are well documented in research studies on the automated detection of hate speech where the highest inter-coder reliability that was reached was 33%[29]. It was only in March 2017 that the Law Commission of India came out with its Report No. 267 on Hate Speech in line with Supreme Court directions on the judgement Pravasi Bhalai Sangthan v. Union of India & Ors., AIR 2014 SC 1591 to examine the issue of hate speech, resolve definitional issues, and make recommendations[30]. Benesch argues hate speech is too broad a conceptual category to understand speech acts that could act as early indicators of translation into actual violence[31];[32]. She identifies 'dangerous speech' as a sub-set of hate speech and a speech category with the capacity to catalyse violence by one group against another (2012).

The Dangerous Speech framework was developed by Susan Benesch when she "noticed striking similarities in the rhetoric that political leaders in many countries have used, during the months and years before major violence broke out". Dangerous Speech is defined as:

Any form of expression (e.g. speech, text, or images) that can increase the risk that its audience will condone or commit violence against members of another group[33].

Violence within this framework is understood to be direct physical or bodily harm inflicted upon people and does not include doxing, incitement to self-harm, or discrimination[34] even though they create the enabling the enabling environment for the violence to occur.

The Dangerous Speech framework attempts to detangle the 'thick concept' of hate speech replete with different meaning and evaluative load as it negotiates social relationships of power and marginalization[35]. It attempts to pare down the negative effects of hate speech to its potential to trigger offline violence and identifies elements that constitutes such speech acts[36]. These include:

1. Message: Dangerous speech deploys the use of coded language in terms familiar to the in-group but not to the out-group, often containing rhetorical patters or shared ideas. It usually contains 5 hallmarks: dehumanization, accusations in a mirror, threats to in-group integrity or purity, assertions of attacks against women and girls, and question in-group loyalty.
2. Audience: Dangerous speech is most effective with a susceptible audience and strategies that build in-group cohesion and collective identity.
3. Context: Social and historical context in which dangerous speech occurs include history of violence and systemic discrimination, competition between groups for resources like land, water etc.
4. Speaker: An influential speaker or authority figure tends to amplify the danger inherent in dangerous speech. However, the speaker need not be an individual but can be organization, group, government, or even a bot. Sometimes, a speaker makes a message not by creating it but by using existing information to re-purpose it through re-contextualisation and re-scripting.
5. Medium: This includes whether it was transmitted in a way that can reach a large audience; involved repetition in its capacity to persuade; use of local language; lack of alternative media etc.

[29]Kwok, I. & Wang, Y. (2013). Locate the hate: Detecting tweets against blacks. AAAI.

[30]Law Commission. (2017). Hate Speech: Report No. 267. Law Commission of India. Available at: http://lawcommissionofindia.nic.in/reports/Report267.pdf

[31]http://lawcommissionofindia.nic.in/reports/Report267.pdf Benesch, S. (2012). Dangerous Speech: A Proposal to Prevent Group Violence. World Policy Institute. Retrieved from https://worldpolicy.org/wp-content/uploads/2016/01/Dangerous-Speech-Guidelines-Benesch-January-2012.pdf

[32]Benesch, S. (2014). Countering Dangerous Speech: New Ideas for Genocide Prevention. Dangerous Speech Project working paper. Washington, DC: Unites States Holocaust Memorial Museum. Retrieved from: https://dangerousspeech.org/countering-dangerous-speech-new-ideas-for- genocide-prevention

[33]Dangerous Speech Project. (04 August 2020). Dangerous speech: A practical guide. Retrieved from https://dangerousspeech.org/guide/ [25 October 2020].

[34]Ibid.

[35]Pohjonen, M. & Udupa, S. (2017). Extreme speech online: An anthropological critique of hate speech debates. International Journal of Communication, 11(2017), pp. 1173 – 1191.

[36]Dangerous Speech Project. (04 August 2020). Dangerous speech: A practical guide. Retrieved from https://dangerousspeech.org/guide/ [25 October 2020].

Pohjonen & Udupa acknowledge that the Dangerous Speech framework recognizes the importance of communicative dynamics in distinguishing dangerous speech from other types of hate speech[37]. However, they argue that being rooted in a global rights discourse leaves limited space for the analysis of cultural dynamics shaping online practices. Adding to this argument – Dangerous Speech presents an indispensable framework for post-fact analysis but overlooks the more processual aspect of growth and mobilization of in-group communities and their pathways to creating an enabling environment for violence to be normalized. Further, it seeks to widen the ambit of violence to include structural violence in form of economic and social boycott.

Pohjonen & Udupa posit 'Extreme Speech' as a form of anthropological qualification in place of the regulatory term of hate speech. Extreme Speech refers to spectrum of practices, which push the boundaries of acceptable norms of public culture toward what the mainstream considers a breach within historically constituted normative orders[38].

Taking forward the call for the analysis of hate speech as practice or practiced speech, this report seeks to incorporate the analysis of cultural elements structuring social relationships. It retains the regulatory scope of hate speech but understands it more as an exercise of structural power shaping social reality – as a process rather than an instantiation. It is related to the historical embeddedness of power, hegemony, and culture of the in-group within the dominant discourse and its capacity to produce the out-group as the oppositional Other through tropes, stereotypes, and rhetorical patterns[39].

Hate speech is directed at a particular group based on their ascriptive characteristics and mobilized by the in-group (the constellation of speakers, audiences, and actors coalescing around a collective identity) which works to reinforce its boundaries to the spatial and discursive exclusion of the Others. Social media creates visibilities or a process of becoming which goes beyond being physically visible as a matter of gaining discursive attention and recognition[40]. Thus, visibility becomes something to be achieved like power, status, and authority[41]. Acquiring visibility also allows then to bestow visibility on certain aspects of the world thereby shaping discourses pertaining to them[42].

In order to bring this into practice the in-group mobilizes meaning – making resources – in the online space this can be related to the different types of content shared. This process of meaning – making is both social structuring and in itself socially structured[43]. This inter-subjective production of meaning involves both listener/ reception as well as speaker/ production who co-construct social action and interaction within the emergent phenomenon of meaning and motive[44]. These serve to create an internal discursive logic within the in-group comprised of actors, social relations, and practical contexts[45].

These interact to produce the overall configuration of social action which are instantiated through calls to action made by the in-group as an enactment of its discursive logic. Such enactment engages social action and social processes with the ways of "thinking, specific identities, emotional responses or commentaries, vocabularies of motives, goals, and reasons for action that are available to the various actors and frame the situation in which the actors 'find' themselves"[46].

"Hate speech is directed at a particular group based on their ascriptive characteristics and mobilized by the in-group (the constellation of speakers, audiences, and actors coalescing around a collective identity) which works to reinforce its boundaries to the spatial and discursive exclusion of the Others."

[37]Pohjonen, M. & Udupa, S. (2017). Extreme speech online: An anthropological critique of hate speech debates. International Journal of Communication, 11(2017), pp. 1173 – 1191.

[38]Ibid.

[39]Saïd, E. (1978). Orientalism. Pantheon Books: United States.

[40]Chow, R. (2010). Postcolonial Visibilities: Questions Inspired by Deleuze's Method. In S. Bignall & P. Patton (Eds.), Deleuze and the Postcolonial, pp. 62 - 77. Edinburgh University Press: Edinburg.

[41]Ibid.

[42]Ibid

[43]Fairclough, N., Jessop, B. & Sayer, A. (2010). Critical Realism and Semiosis. Department of Sociology: Lancaster University.

[44]Ibid.

[45]Ibid.

[46]Ibid.

However, an instantiation of enactment will not suffice to maintain the discursive logic. The stabilization of the discursive logic is required to ensure that the potentiality of future course of action is preserved. The stabilization of the discursive logic will require the establishment of norms an idealized practices through strategies that aim to structure idealized emergent social realities and action.

Further, the maintenance of this discursive order requires the establishment of a sense-making organizational structure in the form of narrative which rationalizes the need of in-group assertion and exclusion of Others. Narratives have 2 functional elements - one is indicative, the other is interpretive[47]. The indicative component serves the function of reportage or description while the interpretive component serves the function of explaining the above description[48]. These two components working together co-constitute meaning[49]. The first step entails the identification of two broad sequences, one of problem definition and the other, of response[50].

Thereafter each sequence so defined is constitutive of and bifurcates into a series of linked micro-sequences classified as per levels of analysis, from generalized (the narrative) to particularized (its component parts)[51]. The singular or combined effect of any number of such micro-sequences represents moments of context dependent risk or the potential to influence and alter linkages to subsequent micro-sequences[52].

This report attempts to understand the phenomenon of practiced speech as a process that works in a continuum to create social contingencies of action. It engages with the question how does hate speech as a practice elicit conditions of violent action.

“This report attempts to understand the phenomenon of practiced speech as a process that works in a continuum to create social contingencies of action. It engages with the question how does hate speech as a practice elicit conditions of violent action.”

[47]Guha, R. (1988). The Prose of Counter-Insurgency. In G. Chakravorty Spivak & R. Guha, Selected Sulaltern Studies. New York, Oxford: Oxford University Press.

[48]Ibid.

[49]Ibid.

[50]Ibid.

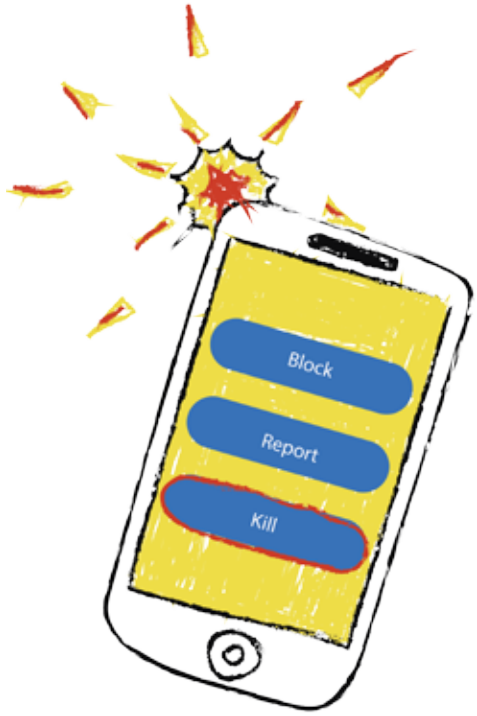
[51]Ibid.

[52]Ibid.

METHODOLOGY

The study employed digital ethnography as a methodology to understand the situated practice of hate speech. This included pre-ethnographic scoping and identification of themes to tap into the broader context of hate speech practices. This pre-ethnographic scoping involved secondary research and informant interviews with stakeholders who have worked on issues of civic violence and hate speech. These formed the basis of keyword searches on Facebook to identify potential groups and pages. This was followed by field research interviews with groups and organizations with assertive in-group identities to understand their positions as producers of content and their engagement as a collective audience in online spaces. The names of pages and groups elicited from the field research was added to the long list of pages and groups that also included the results of key word searches.

Out of this long list only those pages were shortlisted that had at least 5 direct calls to action against a particular community on the basis of identity-based Otherisation. The final short-list included 27 pages and groups. The pre-ethnographic scoping period also involved loose ethnographic scoping across long list of pages to identify and concretise common metrics and emergent themes for recording observations. A common format was decided to capture date of observation, date of publication, period of observation, brief description, content type, strategies, calls to action, engagement (like, love, haha, angry, surprised, tears, loving, comments, and views in case of video). These were complemented by daily observational notes in the form of an ethnographic diary.



> In terms of terminology, this report will use the terms in-group to signify the collective identity that mobilizes functional narrative elements and animates discursive logics into action; and Others to signify groups with ascriptive identities that are (re)produced through representational elements by the in-group to normalize instances of structural and violent exclusion.

The pre-ethnographic scoping period was followed by a four month long ethnographic observation of the short-listed pages along with logging of observations as per format and recording daily notes. At the end of this observation period the data was aggregated and analysed to develop inductive theory and frames of analyses. In terms of terminology, this report will use the terms in-group to signify the collective identity that mobilizes functional narrative elements and animates discursive logics into action; and Others to signify groups with ascriptive identities that are (re)produced through representational elements by the in-group to normalize instances of structural and violent exclusion. The use of such terminology helps to navigate the ethics of online ethnographic observation, prevent the essentialization of identity and practice, and create replicable frameworks of analysis.

MODALITIES OF SPEECH PRACTICES

Speech that seeks to erect boundaries and solidify identities or develop cultures of hate and exclusion can be better understood in the form of practices built upon and building and redefining new structures of social relationships. This highlights the importance of understanding how such speech practices interpellate both the speaker's subjectivity and that of the spoken subject. These subjectivities are then fixed through narrativization of a social reality and their underlying discursive logics. The functional elements of such a narrativization is built through the mobilization of meaning-making resources like content types, enactment of discursive logic through calls to action, and the establishment of the norms and idealized practices through strategies that aim to structure idealized emergent social realities and action.

CONTENT TYPES

7 distinct content types that were observed:

Links to news articles

Links to news article shave significant circulation as it is often used as the means for assigning legitimacy to a particular information. It is often signified as a basis or proof on which perceptions about certain communities are framed or justified. Links to news articles can be divided across 4 types: (a) National broadcast media outlets; (b) Regional news outlets; (c) Online media and opinion channels; (d) Links to miscellaneous websites of organisations and collectives that carry news, views, and updates that are sympathetic to the in-group cause.

However, the news articles that are shared and way in which they are shared have a framing logic. Particularly, links of those news are shared that are exclusive to crimes perpetrated by the Other like theft, rape and even begging. These links are then often observed to be shared with one-liners and/ or with rhetorical questions which often leaves the overall message of the news at a cliff-hanger. Thereby, priming the audience for an pre-frame interpretation which is also inferred from reinforcement of potentially intended message in the comment section. Such news is framed as how the Others are the prime conspirers of vices in the society, most particularly against the in-group and its affiliate members.

This serves to prime the in-group identity towards normalising and justifying violence against the Others. Exclusive attention to the particular framing of the issues and the selection of news articles so shared inhibits responses to others; thereby circumscribing the issues to those which can elicit a collective or individual response[53].

Use of memes and humour

Comics and memes were observed to be shared either directly as posts, or in the comment section as replies to the posts which carry elements of targeting or mocking the perceived opponents. These were targeted at the Others and are instrumentalised to reify the in-group narrative around them. It worked to (a) Reinforce the agenda of social boycott; (b) Reiterate and repeat shared political views and allegiances; and (c) calls for shared advocacy and/ or activism.

Multi-modally shared humour using disparagement helps to forge in-group solidarity and becomes a vehicle for shared meaning and ideology[54]. Production and dissemination of user-generated content in the form of memes, comic strips, sarcastic posts, and humorous content serve as the venue in which certain aesthetics can flourish through linguistic signifiers[55].

Memes are speech acts, for the creation of which certain semiotic or meaning – making resources or signifiers are marshalled[56]. Memes and humour as objects and vehicles for meaning-making work to dehumanise perceived enemies. This is done by building shared practices for the performance of the worldviews shared by the in-group's collective identity. Since such content are precise terms of representation of shared beliefs they are easily digestible and are able to self-propagate through sharing. This leads to high engagement metrics reflected through like, laugh, and love reactions on such posts.

[53]See Park, R. E. (1940). News as a form of knowledge: A chapter in the Sociology of Knowledge. American Journal of Sociology, 45(5), 669-686. Retrieved from https://www.jstor.org/stable/2770043.

[54]54 Dynel, M. (2020). Vigilante disparaging humour at r/IncelTears: Humour as critique of incel ideology. Language & Communication, 74, 1-14. Retrieved from https://www.sciencedirect.com/science/article/pii/S0271530920300410.

[55]Decook, J. R. (2018). Memes and symbolic violence: #proudboys and the use of memes for propaganda and the construction of collective identity. Learning, Media and Technology, 43(4), 485-504, DOI: 10.1080/17439884.2018.1544149.

[56]Grundlingh, L. (2017). Memes as speech acts. Social Semiotics, 28(2), 147-168, DOI: 10.1080/10350330.2017.1303020

Videos

There are overall 5 types of video with discernible purposes that work of social media traction and engagement:

- (Videos by authority figures which serve 'educational' purposes for the in-group with regard to conspiratorial subversion by the Other. These often border on essentialism and use rhetoric and argumentation to demonstrate how it in the Others' very ethos to undermine the in-group which has continued to happen unabated structurally and historically. These are mostly pre-recorded and cross-posted across groups and pages and used to garner engagement and traction. These authority figures have significant online following as well as influential offline presence. These video also become the pivot for mounting campaign and advocacy around a given cause.

- Videos showing state action against the Other are often shared to in order to derive retributive pleasure. These highlight how the Others got what they deserved for being the constant source of disruption of public order in society. These kinds of videos garner reactions in the form of Facebook features like laugh and heart reacts, and applause in the comment section, justifying and glorifying violence exhibited.

- Videos showing instances of oppression by the Others in their sites of cultural and social practice or in contexts where they are in a position in power. These serve to demonstrate how location of their culture and practice are also sites of oppression and abuse. This works to underscore the inseparability of the oppression and abuse from the Other and their very nature thereby justifying and normalising the violence perpetrated against them.

Text posts demonstrate how disaffiliate members of the in-group who share the nominal identity have worked conspiratorially with the Other for structural, historical, and discursive subservience of the former.

- Facebook Lives are real-time videos streamed by individuals often in a planned manner. The agenda and time of holding the live video are announced well in advance. The video usually start with the first 10-15 minutes devoted to salutation, greetings, and audience engagement. This also includes inviting the audience to share the live video on their timelines and/ or groups or use the 'watch party' feature on Facebook that allows individuals to watch videos on Facebook simultaneously in real time. The speakers often wait to start their speech until the video reaches a certain number of shares. Live videos offer a deeper and more interactive engagement in real time. The technical advantage behind the live videos is that they are difficult to be immediately taken down by Facebook unless widely reported.

- Historical videos that show the footages of in-group assertion which are circulated closer to specific contemporary incidents that stand testament to such assertion and 'victory'. This migration of historical videos on to modern technology works to foster a sense of belongingness and mutually recognisable response across wide social distances[57].

Text Posts

Text posts demonstrate how disaffiliate members of the in-group who share the nominal identity have worked conspiratorially with the Other for structural, historical, and discursive subservience of the former. Further, how this alliance has historically eroded ancient cultures and led to societal decay and disintegration.

Such posts are also meant to serve 'educational' purposes for acquainting the audience with narratives of a glorious past which they stand to be denied because of the Other and how such a past needs to be regained. Some posts are long and eloquent which are intended as in-depth analyses and historical narratives often with a few sources at the end aimed at providing credibility to the readers. Such narratives are often shared with the indication about how the in-group has remained so far excluded from their own history.

Apart from long educational posts, posts are often shared with one-liners, poetry with disparaging comments or calls to action, and catchphrases or slogans with the aim of increasing engagement. The purpose of such posts is to make quick

[57]Rajagopal, A. (2001). Politics after television: Religious nationalism and the reshaping of the Indian public. Cambridge: Cambridge University Press.

announcements or keep the followers or group members engaged with pages/groups/profiles, because often when a person stops posting for a day or two then the comments on the newest posts start pouring enquiring after the person. So a quick salutary message becomes a way to 'sign-in' and mark their presence on social media.

Posters, infographics, and screenshots

Poster and Infographics are used as objects and vehicles of information that are curated within the ecosystem of these pages and groups. The information that is intended to deliver can be further classified into two types:

(a) Detailed infographics where elaborative timelines or chronology of historical events, mythologies, comparative cultural practices of the in-group and Other, discussion of political ideologies, and contemporary news are curated and shared.

(b) Posters that are agenda-driven are precise and involve calls to action, metaphorical comparisons, and modes of campaigning and advocacy.

Screenshots of tweets, news, and photographs of key public figures are often shared. Screenshots also include crimes committed by the Other apart from photographs of public figures admired by the in-group. However, the veracity of these screenshots are difficult to determine and in certain instances at least the news shared therein have been debunked by fact-checking organisations.

Multi-platform linkage

A common practice involves sharing links to Twitter, Telegram, Instagram and YouTube accounts in order to expand their audience and reach. It is shared with the captions urging followers to mark their presence on other platforms as well. The ability to maintain multi-platform presence has been utilized by online actors including public figures whose primary presence is rooted outside social media. There have also been instances of fostering public participation by encouraging audiences to participate in Twitter polls to determine a public figures' topic of discussion. There are also online influencers who rely primarily on social media for their popularity, and therefore exist across platforms to generate wider audiences for their content.

"Call to action is to urge the followers or the audience to carry out certain specific actions against the Other. Out of the 7944 posts observed, 1898 (24%) contained calls to actions either in the form of (i) call for direct violent action; (ii) call for economic boycott; (iii) call for social boycott; (iv) call for extreme action by the government."

Self-aggrandisement

Social media influencers, activists, and leaders often used the method of uploading their own photographs and captioning them with assertive and disparaging text against the Other. Apart from this, individuals post their pictures on public pages with swords and guns in order to display muscle power.

CALLS TO ACTION

Call to action is to urge the followers or the audience to carry out certain specific actions against the Other. Out of the 7944 posts observed, 1898 (24%) contained calls to actions either in the form of (i) call for direct violent action; (ii) call for economic boycott; (iii) call for social boycott; (iv) call for extreme action by the government. Same posts can have a combination of one or more call to action. However, it was also noticed that those who regularly post content are potentially wary about inserting calls to action in their posts, since if over-used, it may lead to unwanted attention and reporting and shutting down of the account.

The call to actions demonstrate that such actions, when self-perpetrated, become justified because of the perceived crime or wrongdoing from the Other side as a result of framing of the information shared on these platforms. Herein, the division of Self and Other is rendered extremely unequivocal. Through this process it becomes easier to demand the most extreme action towards the Other through the idea of retributive justice. Where the Self becomes the arbiter or reinforcer of justice and their enactment, thereby ascribing to the Other a low level of morality are moral righteousness to the Self.

The calls to action seemingly come with an urgency of an immediate battle which needs to be handled right at the moment. While calls to action can potentially be used as a means of holding sway over the audience and retaining their engagement; these also have serious potential to culminate into offline violence and action as they becomes a means to

mobilised and gather crowds to designated locations and carry out intended actions.

Calls to action usually stem from the desire to correct the historical wrong that is still ongoing. The justification for the current call comes from historical and contemporary 'evidence' provided through various content types mobilised through strategies. Strategies (described below) are of significant importance underpinning calls to action since strategies, over time, create stereotypes and justifications for call to action. Call to actions claim their need and essence from the stereotypes, accusations and justification built over time to the point where perceptions about Others are reified based on their identity as well as the rationale for some sort of retaliatory action.

Historically, as was seen in the case of Rwanda and Germany, speech acts have had immense potential to galvanize existing prejudices to the point of taking action or normalising violence. Incitement of violence through social media platforms is both a process as well as an event. As a process it involves the systematic Otherization of a particular community or group of people to the point where asking for/ resorting to violence against them is considered justifiable.

Cultures transmit norms and beliefs across generations through observation and imitation. Norms and practices culturally encoded as desirable or necessary lead to the creation of an enabling environment for violence to occur, be normalised, and justified. Examining the socio-cultural and historical undertones behind the kinds of direct actions that were called for within the digital communities that formed our site of research, we find that they can almost always be tied back to a common larger picture, a common goal or the overarching narrative – that of in-group assertion and spatial-territorial claims to the extent of violent/ structural exclusion of the Others.

There were 4 types of calls to action that could be observed:

(a) Call for direct violent action:

A call for direct violent action is directed against a particular community or an individual based on their beliefs, cultural, or social position. The call for violent action is not just restricted to physical violence in the form of beating, lynching, or killing; it also includes calls for sexual violence against women and anyone in general that shares the common identifier.

Out of 1898 posts containing calls to action, close to 34% (637) contained a call for direct violent action. It was observed that there were instances where individuals posted statuses about their accounts being reported for the content shared by them and a few of them were even suspended.

As a result of this, there is greater focus on the utilisation of multiple strategies as a longer play to set ideological narratives, particularly by influencers, authority figures and opinion leaders. However, direct calls to action are observed to increase in occurrence in response to particular event – particularly those involving public crises.

Direct call to action is posted as a form of retaliation against direct offence, as a mark of retributive justice, or a part of self-defence or self-preservation. Direct call to action is also about interpreting the call since it may not be as direct as it seems in order to avoid any legal problems. A lot of times call to action comes from the audience because of certain posting strategies in which the caption of the posts engage the audience by asking for opinions and suggestions about what to do.

It invokes the idea of taking matters into (one's) hands; therein showing lack of faith in the political and judicial system. This also includes the call for cultural reversals through violent means and destruction or demolition of their sites of cultural, social, and religious practice.

Such calls are often justified as retribution based on the narrative of historical oppression and contemporary conspiratorial subversion done by the Other. It is presented in a way that the said incident is in the pattern of a historical continuum; thus, the moment has arrived for a violent reply.

The demand for violent action is also created strategically in a way that primes the audience through instigating questions in captions as a part of

"Out of 1898 posts containing calls to action, close to 34% (637) contained a call for direct violent action. It was observed that there were instances where individuals posted statuses about their accounts being reported for the content shared by them and a few of them were even suspended."

a multi-modal post. This is often posed as rhetorical questions asking for a response from the audience in conjunction with presenting information in a disparaging and dehumanised way. This capitalises on the insecurity created through false information which then leads to building up a sense of emergency where an action is required.

Call for violence is also often done through sarcastic memes which rather than being explicit, call for violence by using humour as a means to prime the audiences and reify shared beliefs and insults. Sometimes skilled couplets are used as dehumanising strategies which are then capped off a violent call to action.

Another prominent method is usually through live videos, where the person delivering the live video could be seen as visibly angry or with tears in his eyes; thus, channelling and transmitting shared emotions among the in-group audience. This helps to build solidarity and coalesce around the ideas of a shared hurt or insult. Live videos broadcasted almost in real-time to followers and audiences, allows them to engage in forms of interaction with the speaker.

A live video cannot be immediately taken down by Facebook unless reported widely. It can only be taken down after it is being reported enough number of times by the users. However, given the live videos are often directed at the broadcaster's intended audiences, by the time enough people find the content objectionable it will already have been widely in circulation.

Live videos also act as a platform of delivering speech which are not just delivered as a monologue; rather, as a fiery enactment of topics and issues circulating within the online spaces inhabited by the in-group. The more furious this enactment is, the higher the chances are for the speaker to give any call to action, in order to create a lasting impression on the audience.

Calls to action are often enmeshed with sexually coded speech both in the form of speech as usual, to underscore a point, or in the form of humour. Captions use sexual slangs for making fun as well as as a threat against the opposition. Normally the captions are posted to glorify themselves or their leaders and tend to use extreme sexual slangs to reflect dominance.

However, calls to action also include calls towards direct sexual violence in the form of rape or physical assault especially against women targets who are threatened with rape or other forms of sexual violence. Most abuses carry rape threats or threats of sexual assault to both women and men with threats such as using weapons to hurt or maim their genitals. Even the violence demanded through state authorities also includes sexual violence of a similar nature.

Calls to action also relate to demolition and destruction of sites of cultural or religious significance as a means to claims-staking layered with retributive justice. Call to violence in the form of warning of retaliation is something that is used often which is why the word 'if' becomes important i.e. if a particular action is taken by the Other it would be met with violent action that is being called for. A lot of calls to violence are often murky with statements that the consequences of given acts by the Other will 'not be good'.

There are several videos of offline rallies or demonstrations where either the speaker's words or the crowd's chants are calling for violence. These speakers and gatherings also intersect across groups/ pages, giving us a glimpse at the kind of associations of solidarity that exist offline.

(b) Call for economic boycott:

Calls for economic boycott amount to 21% (397) of the total number of calls to action (1898). Economic boycotts usually occur in tandem with an issue or a controversy. This is also followed up with offline distribution of signifiers like flags which acceptable business owners can display with their businesses so that they can be easily identified by the in-group and prevent further harassment.

For this type of call to action, there comes requirement to prove how there is an underlying conspiracy to get economic benefit out of the in-group through systematic '(mis)allocation' of resources through the argument that Others are allowed to have their own economy based on their social and cultural practices which is denied to the in-group as a result of systematic erosion, subversion, and suppression of their cutlures.

However, it is argued that such an economy is only able to survive with support beyond their community

"Calls to action are often enmeshed with sexually coded speech both in the form of speech as usual, to underscore a point, or in the form of humour. Captions use sexual slangs for making fun as well as as a threat against the opposition."

which includes uninformed members of the in-group which form its largest consumer group. It is further argued, that it is this economic power that end up marginalising the in-group and prevents it from staking its legitimate claims.

The boycott calls are based on the argument that if the in-group stops taking part in Others economy and instead works for fostering their own economic arrangements, it will make the in-group stronger. Such posts mention calls for boycott of the Other as well as disaffiliate members of the in-group who aim for more co-operative social processes. Another method deployed is to show alleged scientific or health hazard of using products sold by certain communities.

Often, the mode of advocating a call to action was to urge to have economic relations within the community. This does not directly mention boycotting of Other communities but promotes economic interaction on the basis of in-group identity. The purported aim is to have the in-group community to be as close-knit as the Others are alleged to have. However, this not only relates to the in-group in general but also works to reinforce sub-group affiliations.

The call to action for economic boycott is not just restricted to business establishments owned by the Others but also extends to transgressions committed by public figures with a call to boycott their work. The aim is to make their work commercially unviable and demonstrate the consumer power of an united in-group which can influence major economic decisions.

(c) Call for social boycott:

This involves asking people to be not allowed on public or private premises based on their identity; this also extends to cutting off social ties or maintaining social distance beyond grounds of public health. Calls for social boycott amounted to 12% (225) of the total calls to action (1898).

The first form of social boycott is aimed at self-preservation of the in-group with Others defined as being harmful to health, security, and well-being as a result of their practices and proclivities. This extends beyond individual social boycott to calls for institutional social boycott so that the overall populace can be safeguarded and protected. Health threat is an easy topic to build an argument on given the ease with which health scares can be provoked among general populace. Here too calls to action come in the form of humour tying in with contemporary events and public announcements and practices.

The other form of social boycott is applied against disaffiliate members of the in-group community who appear to be weakening in-group ties by promoting inter-group solidarity. These disaffiliate members are often seen as a part of the larger conspiracy to undermine in-group claims-staking. The aim behind such calls is potentially to reduce the influence of such key disaffiliate members on in-group solidarity through a disparagement or negation of their work.

The third form of social boycott relates to boycott of inter-group relations as this purportedly leads the members of the in-group astray from their true cultural and social path and practices in the interest of relativism. This also included regulation of women so that the Other cannot make incursions into their communities through love or marriage alliances. It is argued that it the Man's responsibility to ensure social boycott of the Other in order to protect their women.

Like other calls to action, these gather momentum ahead of and during certain epochal events or events of public crises. In advocating for in-group purity, there are also moves to advocate for sub-group purity of cultural and social relations as a step towards protecting their cultures from erosion through mixing with others.

There are posts that urge people to give work and provide help to their own communities, rather than helping without identification. There are advertisements or helplines for financial help and jobs posted for people from specific community which is not only a form of social boycott but also violation of fundamental rights given by the constitution since it does not allow discrimination on the basis of identity.

Another aspect of social boycott involves uniting the in-group community for offline action against cultural and social practices that are not their own but operate in shared public spaces. News, videos, and examples of these surface as exemplars of actions that initiate claims-staking and serve as the instantiation of success in enforcement of local regulatory norms while also advocating for institutional banning of such practices.

> The first form of social boycott is aimed at self-preservation of the in-group with Others defined as being harmful to health, security, and well-being as a result of their practices and proclivities.

(d) Call for extreme action by the government:

Calls to action also involve asking state authorities to take extreme measures against people from a certain community. Under this category, calls to action include (i) Change in law or constitution in a way that can right 'historical wrongs'; (ii) Direct calls for violent action from the state authorities. This highlights the multiplicity of approaches, all ultimately aiming towards cementing spatial-territorial claims-staking by the in-group.

It was observed that there are those who would like to legislate such changes into existence through a gradual and more pervasive process while there are others who feel a more pressing need for immediate violent repression towards Others.

The idea articulated most often is a call for state action for institutional acceptance of spatial-territorial claims and designate such officially through legal and constitutional changes. It is argued that the disaffiliate members of the in-group help Others continue their domination over the in-group thereby aiding in their continued subjugation. This has happened in a historical continuum which has caused the in-group to lose their cultural and social identity.

Calls for state action also extend to institutional banning of cultural and religious practices that take place within shared public spaces and affect in-group beliefs and sensibilities. This also extends to symbolic practice of claims-staking and assertion of identity through affiliative symbolism.

Specifically, around events of public crises the call for violent state action escalates to calls for extreme violence against the Others who are perceived to be the harbingers of the given crises. This is done as a part of shared and felt retributive justice. This demand for state action is not just in the form of physical assault or beating but also demands for sexual assault with notions of penetration and assault on genitals. While there is demand for maiming of genitals in the form of cutting off or burning for sexual crimes, the demands for perpetration of sexual violence against women targets are fierce.

Calls for state action also extend to cultural regulation with bans on literature, art, and cinema that are violative of the in-groups norms, beliefs, and practices. Demand for stricter laws for the protection of such norms, beliefs, and practices with greater punitive actions against violations and those responsible for it. This is supplemented by multi-modal examples of the in-groups' offline extra-judicial regulation of such.

STRATEGIES

Observations showed coherent patterns of narrativization, rhetoric, information dissemination as well as response. Towards reifying in-group solidarities, identities, and cohesion common modalities of actions observable across and within the online networked take the shape of strategies. These involve determining certain actions and mobilising discursive meaning-making resources within recognisable forms of implementation that work to priming audience identities, particularly that of the in-group.

These leverage embedded networked subjectivities utilising affordances of virality and instantaneous spread within core and affiliated networks and followers of influencers within the network. The observations led to identification of the following strategies utilised to maintain in-group cohesion and solidarity through continuous priming of in-group identities.

Mis/disinformation practices

In India, the permeation of digital technologies in society has primarily taken place through cell phone usage, which has led to private messaging apps such as WhatsApp becoming infrastructural to social interaction across the country. Facebook as a social media platform exists in an inextricable ecosystem with messaging application.

These become both sites of information sharing as well as forging networked connections. Such practices coalescing around one's social media presence as infrastructural to the social fabric of community bonds across geographies in order to form and maintain networked subjectivities.

Misinformation and rumours crop up simultaneously across different platforms/ groups/ pages at the same time. They tend to come loaded with an internally consistent logic justifying the means of their existence and are neither random nor coincidental in their messaging.

An important characteristic of misinformation and disinformation is that they are very much in tune with whatever is unfolding in the daily news cycle[58]. It was observed during this study that several of the infographics and explainer videos (monologues on a particular theme, with audio-visual aid) market themselves as providers of the unbiased truth that is not covered by more established 'mainstream' sources of information.

[58]Informant interview.

Content creators often frame the messages from the standpoint of a historically marginalised segment who must now assert themselves to overthrow the structural oppression. Thereby, creating sympathy and solidarity among the viewer and underscoring the value of their support in terms of visible online engagement.

Narrativised misinformation works to create an environment of reified in-group identity within which there is neither space for or acceptance of the Other. This serves to normalise violent calls to action when they coalesce around events.

This translates to development of hashtags as meaning-making objects which are mobilised into campaigns online and at certain instances result in offline action. These revolve around conspiracies around spatial claims, encroachment of family values, and the Others as harbingers of public crises that endangers in-group lives.

These narrativized disinformation campaigns draw this conspiratorial undermining on a historical continuum of the oppression suffered by the in-group. These narrativized accounts are enmeshed with ideas, tropes, messages, and stereotypes which circulate more widely in the public domain and mainstream media channels[59].

This transmediality within which such narrative accounts unfold on multiple media platforms leads to each account making a distinct contribution to the overall narrative by acting as one of its functional units of meaning-making[60].

Narrativised misinformation works to create an environment of reified in-group identity within which there is neither space for or acceptance of the Other. This serves to normalise violent calls to action when they coalesce around events.

Disinformation campaigns using disparaging speech peak around public mobilisation around civic issues pertaining to the Other in order to delegitimise such issues and civic participation. It also involves wilful manipulation of authentic but unclear video content which is translated in explanation with the accompanied text and rhetoric.

From within narrativized disinformation campaigns two rhetorical strands can be observed: one is based on spatial-territorial idealism as an emergent idea and practice where in-group claims and assertions are realised; the other is based on how the in-group is still in danger from conniving and conspiring Others who have historically used all strategies to ensure subservience of the in-group.

Narrativised disinformation campaigns can play a substantial role in radicalizing people to disseminate threatening or violent speech, as well as carrying out offline action.

Glorification of assertive action

This involves glorifying violence perpetrated against the Others towards normalising the use of violent means as a mode of assertion. This include: (a) glorification of incidents of retributive justice; (b) glorification of structural violence or institutional changes that they deem to be in favour of their spatial-territorial assertion; (c) glorifying public figures – both historical and contemporary that have been symbolic of enacting in-group assertion and mobilising their beliefs; (d) glorifying extra-territorial claims-staking towards the ideal homeland; and (e) mass mobilisation of the in-group in order to regain cultural superiority.

Dehumanisation

Dehumanising tactics or subjectivizing Others as less than human legitimises violence and increases motivation for violent actions[61]. The strategy of using dehumanising metaphors is with the aim to justify agendas or narratives[62]. The family of dehumanising metaphors evokes hostility, disdain, loathing, physical disgust, and/ or bodily fear in people[63].

these metaphors simultaneously dehumanize their targets and justify the repressive and inhumane actions that are taken against them. Indeed, they

[59]An observation also recorded in Banaji, S., Bhat, R., Agarwal, A., Passanha, N., Pravin, M. S. (2019). WhatsApp vigilantes: An exploration of citizen reception and circulation of WhatsApp misinformation linked to mob violence in India. Department of Media and Communications, London School of Economics: London.

[60]Ibid.

[61]Wahlström, M., Tömberg, A., Ekbrand, H. (2020). Dynamics of violent and dehumanizing rhetoric in far-right social media. New Media and Society, 1-22. DOI: 10.1177/1461444820952795.

[62]Szilagyi, A. (08 March 2018). Dangerous metaphor: How dehumanising rhetoric works. Dangerous Speech Project. Retrieved from https://dangerousspeech.org/dangerous-metaphors-how-dehumanizing-rhetoric-works/ [19 October 2020].

[63]Ibid.

present the hostility, policy restrictions, maltreatment, human rights violations, and physical aggression to which those people targeted are often subjected to as necessary and that can be carried out according to bureaucratic procedures — naturally excluding any emotional identification with the victims[64].

The dehumanisation strategy was most evident in: (a) use of profanities and slurs to describe the Others; (b) use of metaphors of vermin, pigs, snakes, cockroaches, pests; (c) framing of cultural, social and/ religious practices and sites as oppressive, backward, or being the incubator of violent action; (d) labelling Others as foreigners and invaders; and (e) either the hyper-sexualisation of women targets or painting them as perennially voiceless victims.

Constructing hate typically requires the invoking of an allegedly predominant identity that drowns out other affiliations. The removal of a category of people from one's moral universe by categorizing them as sub-human, is a key to the link between dangerous speech and physical harm. It is a part of a process of moral exclusion of the Other in the collective consciousness, a process that also includes and calls upon political and legal institutions to legitimize its messaging.

Given that similar dehumanising language is being used by a wide range of unconnected actors in the network, it has become a part of the shared vocabulary. It potentially underscores how as a strategy it can be used to prime in-group identities.

Stereotyping

Stereotypes offer reductive and essentialist metaphors, myths, or beliefs that categorise some humans as 'normal' and the Others as 'abnormal'.[65] This allows one group the power to represent, constrain, exclude, and punish those defined as the Other[66]. Through repetition and reassertion in cultural environments these stereotypes come to seem natural and timeless even to those subjugated by it[67].

Stereotypes included designating the Others as harbingers of public crises and spreaders of disease. This is enmeshed with paranoia about practices which were translated as the modes of aggravating of crises of transmission of disease – a threat to public health, particularly that of the in-group. Stereotyping gets mobilised into campaigns which then spirals into public action when violence is perpetrated against members of the Other community. This escalates in the backdrop of public crises.

Against the backdrop of events of public crises, stereotyping transmutes to scapegoating mobilised through viral campaigns which get aggravated at first due the selfishness of the Other moving on to a larger conspiratorial plot devised to undermine and subjugate the in-group.

The power of this narrativized mobilisation is such that even videos showing the most innocuous things like a member of the in-group merely touching something gets translated as evidenced of a wider conspiratorial plot and subversive practices. This leads to collective panic and prejudice forming a vortex in which violence is justified whether they be in the form of physical assault, police brutality, denial of medical treatment or economic and social boycott.

Video depicting attempted rape of a minor and ensuing mob justice meted to the perpetrator is translated as sexual depravity being an universal attribute of the males in the Other community. Similarly, there are depictions of mob justice being meted out in the instance of inter-community love affairs. Such instances are used to highlight how the in-group's family life and social fabric stands to be 'infiltrated' through marriage alliances and loves affairs.

This is reiterated and repeated by public and authority figures on how looting, violence, murder, and subjugation are essentialised and inscribed into the very socialisation of the Other. And it is argued that it is this very essentialised nature with which they would prey upon the female members – sisters and daughters – of the in-group should its men not be able to violently defend its social boundaries.

These often get translated to harassment and mob-justice on the streets through cornering stray individuals with blows and identity-based slurs. This gets translated into rants about how they would bring downfall of all the social order and how every trace of their cultural and social identity should be razed to the ground and obliterated and how, in one instance at least, the perpetrator claims to be ready to bathe every member of the community with acid.

There is persistent messaging aimed at the creation of a homogenous perception of the Other as having unfavourable attributes. This includes the idea that they are 'unclean', are lewd and overtly sexual (the former is particularly used for men), and that they are uneducated and backward. Two of the most popular

[64]Ibid.

[65]Banaji, S. (2017). Racism and orientalism: Role of media. In International Encyclopaedia of Media Effects(2017). Wiley-Blackwell ICA: Online. DOI: 10.1002/9781118783764.

[66]Hall, S. (1997). The Spectacle of the Other. In S. Hall. (1997). Representation: cultural representations and signifying practices, pp. 223-279. London: Sage.

[67]Banaji, S. (2017). Racism and orientalism: Role of media. In International Encyclopaedia of Media Effects (2017). Wiley-Blackwell ICA: Online.DOI: 10.1002/9781118783764.

tropes of stereotyping are the tropes of the 'predatory man' and the 'oppressed woman' that needs rescuing.

Social media platforms uniquely enable the scale of targeted dissemination through the re-contextualised older content, or the sharing of violent/ explicit audio-visual content. This is used in the nature of furnishing 'proof' or 'evidence' of realising the stereotype. This enables the manifestation of stereotypes in 'reality' so that it can be mobilised as an accurate descriptor rather than as an instrument of prejudice.

Live streaming

Livestreaming forms an efficient form of attention hacking. The Facebook Live feature mixes the topographies of a live broadcast with the advantages of decentralized and user-driven web 2.0. There is limited editorial oversight and interventions by the platform are contingent on reporting.

Once a Facebook Live video is created, it resides permanently on a page or profile for viewers who missed the live event to view at any point. Videos are eligible to show up in friends or followers' news feed during the live event, as well as after the event has ended.

When Facebook Live is used to livestream an ongoing event, as compared to just sharing a normal video of the same event, the main difference is that livestreams are timestamped, creating a perception of a 'trustworthy' media text that (a) has not been tampered with and (b) allows the audience to experience, even participate through comment sections, in an event in real-time.

Particular uses of live-streaming that were observed:

When Facebook Live is used to livestream an ongoing event, as compared to just sharing a normal video of the same event, the main difference is that livestreams are timestamped, creating a perception of a 'trustworthy' media text that (a) has not been tampered with and (b) allows the audience to experience, even participate through comment sections, in an event in real-time.

(a) Immortalising specific segments from daily broadcast news:

Television news channels have taken to expanding their reach through livestreaming their news segments. These are originally posted by the official Facebook page of the channel in question, and then shared widely by their regular viewers (which are significant in number). Keeping in mind the additional advantage of being able to download, crop and edit Facebook Live videos, the potential for segments that are agenda driven to capture public imagination and form collective consciousness is increased manifold.

(b) Use by individual 'activists' and local organizations with offline presence to mobilise around a particular issue

This involves the given individuals or members of local organisations indulging in action while keeping up a running commentary for the benefit of the audience. This serves as an enactment and translation of the narrative into action. The commentary is often interspersed with slurs and stereotypes along with how members of the in-group need to 'wake up' and find their conscience to work towards the emergent idea of a new homeland. To an in-group viewer such videos demonstrate that these individuals are fighting on the frontlines of a larger battle that they have a stake in. To non-in-group viewer, this video serves as a warning, a precursor for the retaliation that is inevitable for non-conforming.

(c) Used by individuals and public figures to incite violence

The enact of violence is often accompanied by Facebook Live in order to record the act of service to the cause. This is also used as a platform for warning that consequences will follow perceived transgression and includes a call to followers to gather at designated sites and locations in order to implement violent action. Some use this to present evidence in favour of stereotypical tropes and metaphors that are a part of in-group parlance. It also becomes a platform to call for action and mobilise narrativized disinformation.

(d) Use by social or political commentators

This includes commentary on social or political issues – sometimes with individuals speaking while at other times maintaining anonymity through using voiceovers. These are intended as primers of information on current or historical events. However, they tend to use dehumanizing speech, glorifying

violence against the Other, or used narrativized disinformation. Most of them also maintain a multi-platform presence with YouTube channels and/or Twitter accounts under the same name.

Existing mechanisms for intervention by the platform itself are contingent on there being an immediate, coordinated response by an audience watching an individual/group going live on Facebook. This means a large number of people must be flagging the live video while it is live. According to Facebook's Help page: "Depending on the severity of the situation, we [Facebook] may end the live video, disable the account, and/or contact law enforcement."

It was observed that sometime, mass reporting would indeed lead to the live video being cut off mid-sentence but there were similar videos which received no response and continued to remain on the platform. At the times, the given individual might lose access to their account for a couple of days but is able to maintain their presence through a secondary account. Video take downs and blockage of access did not appear to deter individuals from their existing practices.

Creating visual archives of proof

A use of social media platforms is instrumentalized to structure public memory. Speech acts and practices combined with technological materialities of the platforms create a sensory-technical infrastructure of possibility of thought and experience[68]. This offers not just an avenue to guide engagement with something that is unfolding in real time but also lasting archives that structure atemporal aspects of perception and memory.

These include sharing videos of alleged perpetration of violence and assault by the Other, alleged victim testimonies, instances of violent inter-group clashes, pictures of Others carrying weapons, swords, and knives which speaks to their essentialised violent nature, images of grievously hurt or traumatized individuals who are said to have suffered their fate at the hands of the Other. In all of these visuals, the perpetrator is overarchingly the designated Other. This occurs in tandem with voyeuristic gaze trained on targeted women where the speaker offers anyone who cares to ask – nude or sexually explicit videos of such women.

A media upload on the platform remains in the media or gallery sections of these groups or pages, which can be accessed anytime by anyone and be downloaded and repurposed forever. This is also an effective way to create and manage permanent digital archives of both the past and the present: curate visual 'proof' that certain things happened in a certain way irrespective of veracity of such information. A powerful example of this is the ability to 're-script', recontextualise, and repurpose events.

Visual media uploaded online are 'remixed' to annotate an 'event' which exist in entanglements with dominant discourses surrounding that event. Different actors – such as news media, social media, citizens visually compose these events differently[69]. The sharing of graphic content is most observable on private groups since the chances of them being reported remain low.

Internal policing

Internal regulation is often used to maintain the boundaries of the in-group. This translates to declarations of violence against disaffiliate members aiming for cultural heterogeneity. In one instance, the only consequence for a site of cultural assimilation is for it to be burnt down. This is because such sites are seen as a deliberate dilution of culture and a part of the wider conspiracy by the Other to obliterate the in-group from existence.

This relates to intimidation of public personalities attempting cultural heterogeneity. When such personalities retract and apologise for their previous conduct, it is deemed that campaigns for cultural purity have been successful. Apart from online campaigns it also involves offline intimidation of individuals who have made a joke that have affronted the in-groups cultural sensibilities – in one instance a comedian's face was blackened with ink.

Neologisms

Neologisms and puns are a way of developing a shared vocabulary that work to increase in-group cohesion. The feedback loops between mainstream media and social media generate and normalize a shared understanding with its implicit assumptions and shared foundational narratives.

These neologisms are often hashtagged and used as a shorthand as functional narrative elements and mobilised as meaning-making resources to reify in-group identity and cohesion. These neologisms often take the form of puns of existing names, portmanteaus, or sarcastic epithets.

[68]Hirschkind, C. & Larkin, B.(2008). Introduction. Media and the political forms of religion. Social Text, 26 (3), 1–9. DOI: 10.1215/01642472-2000-001.

[69]Sengupta, S. (2013). 'The 'Terrorist' and the Screen: After Images of the Batla House 'Encounter'.' In R. Sundaram (Ed.), No limits: Media Studies from India, edited by Ravi Sundaram, pp. 300-26. Delhi: Oxford University Press.

FRAMEWORK FOR SOCIAL PROCESSES OF DIRECT ACTION

Collective identity building

The identity building of the 'collective self' is a continuous process of assembling dispersed actors on a platform like Facebook by engaging them around a functional narrative mobilised through its component elements in the form of content types, calls to action, and strategies. This identity building process is based upon forming homogeneous categories, signifiers, and signifying action of such collective identity, on the basis of which future courses of actions could be called for.

Collective identities are built and in-group boundaries are reified through interpellation – a process of hailing – classifying, sorting, and assimilating individuals by addressing them with identity markers[70]. It is a process, which, through the affordances of social media platforms are used to reproduce active subjects mobilized through the apparatus of their group identity.

This was observed to happen through religious identity-based salutations on posts during the any given day. This included using religious identifiers of brotherhood to call to action to unite against the enemy 'Other', using cries of religious symbolism to celebrate acts of humiliation of the Other, and perceived victory of members of the in-group.

> The collective identity building is like a solution to the existing moment of crisis brought upon the members of the in-group through a combination of historical subjugation by the Other and their more contemporary entrenchment in society and polity to the collective disadvantage of the in-group; whom 'Other' has otherised in their 'own' territory – signifying usurpation.

These mode of address or phrasal turns are accompanied wherever there are alleged instances of news of infliction of atrocities over the in-group. The call acts as a reminder and a way to assemble the dispersed actors. This necessarily does not always mean a call to pick up arms or mobilise with violent intent but rather acts as a priming for action.

Greetings and modes of address among the in-group often ascribe a spatial-territorial ownership thereby calcifying primordial associations and identities. These statements recur throughout posts and especially in the live videos which act to keep the followers/participants bonded together as it gets repeatedly asked to be commented as a marker of attendance.

The conflation of identity with territorial boundaries excludes the Other occupying the same space. This discursive exclusion leads to the creation of an social vehicle through which they can be excluded through violent means. Further, identity-based salutations are valorized by likening them with symbolism of aggressive bravery like tigers or lions.

Herein, aggression is justified and transmuted to bravery in doing what is required to preserve and secure the in-group against incursions, invasions, and injunctions from the Others. The in-group is thus at once vulnerable to injustices from the Others and aggressively brave to defend itself from them.

Further, the in-group boundaries further differentiates itself from its disaffiliate members. These members often become the subject of in-group derision, at times they are invoked to pontificate the existing or growing power of the 'Other' which brought the moment of crisis where the in-group must mobilise with new found consciousness.

The collective identity building is like a solution to the existing moment of crisis brought upon the members of the in-group through a combination of historical subjugation by the Other and their more contemporary entrenchment in society and polity to the collective disadvantage of the in-group; whom 'Other' has otherised in their 'own' territory – signifying usurpation. In this process of solutionising, the in-group develops a narrative that creates new hybrid subjectivities in which the 'Other' is a product of historical bastardisation – a taint they carry not only in their existence but which has been essentialized in their very nature; and which has been historically instrumentalized to the detriment of the in-group.

[70]Althusser, L. (1971). Ideology and ideological state apparatuses: Notes towards an investigation. Monthly Review Press.

While this narrativization creates the conditions of discursive exclusion of the Other, it privileges the view of the historical fortitude and resistance displayed by the in-group in standing up to their oppressors and how such fortitude is the call of the hour in defending the in-group against similar onslaughts and usurpations. This gets embedded into the process of identity formation in the form of refrain pervading throughout methods of engagement on pages and groups on the platform.

However, the in-group also works to keep its membership fluid through the elision of social-historical divisions within nominal identity of the in-group, discarding previous animosities and using argumentation to 'debunk' popular conceptions underlying the existing division between the members' nominal identities. The seeming projection of an integrative identity that posits an all-encompassing idealism that promises openness and camaraderie to those who are able to subscribe to the views of the in-group.

Despite such tactics of integration, perceived transgressions of the historically marginalized groups within the nominal identity receive admonishment as not deserving of being a part of the civilizing integrative project undertaken by the in-group. It must be noted that while some conversations in some groups were centred around the project of integration, other focused on how the general population of the in-group has been disadvantaged by policies aimed at the social and economic integration of the marginalized sections that has given them a perceived unfair advantage. Thereby, mythologizing the notion of a vulnerable oppressed marginalized group within the nominal identity on the basis of their their perceived aggressive politics.

These subtle contradictions highlight integrative practices so that the in-group can present itself as one. So that certain actions against their perceived enemy can seem to have been decided collectively. It shows that apart from emboldening the outer boundaries of 'us versus them', it is equally important to blur the intra-group lines present within the given nominal identity, for what is deduced from the observations of the pages and groups, there is a perception that 'Others' has always tried to take advantages of these differences by politicising them.

However, this homogenizing project of the in-group is not closed and atomistic but assimilative. It works towards discursively building avenues and interfaces for other groups to merge within its crafting of the space-identity historical continuum where the in-group identity acts as a fulcrum with interfaces through which newer connections can fused with that of the in-group while the in-groups' core identity remains predominant.

While creating and establishing its boundaries, the in-group is also conscious of preserving and policing it. This is done by reminding members of the primodial identities and glorious past including strategies that involve shaming members for being ignorant about this past; and why other members of the nominal identity group remain out of the fold. This is often instrumentalized through deconstructing aspects of popular culture on how it further erodes the nominal identity through its irreverent cultural tropes. This necessitates the integration with in-group is vital to take forward the ideal way of life which was emblematic of the glory of the past. The in-group's claim of superiority of their identity is used to 'awaken' their members from diverted ways of life. Closely knitted in-groups are easier to administer and mobilize against perceived injustices.

This involves a derision of the politics of difference, diversity, and inclusion which are said to undermine the cohesion of the nominal identity with the in-group. These are used to further subjugate collective interests and work towards a placatory attitude that places 'Others' interest before the interest of the dominant group.

The building of the collective identity is a process that is aimed with both developing the collective identity as well as maintaining its clarity of vision and boundaries. This is a daily continuous process that uses multiple strategies and content types. It strengthens the in-group boundaries through highlighting how transgressions against nominal identity goes unnoticed while 'Others' have managed to colonise spaces that should ideally belong to its rightful inhabitants. This necessitates the particular politics practiced by the in-group and rationalizes the requirement of its existence.

Narratives of blame

The creation of an Other is an important factor to differentiate from the collective identity of Self. Mobilising insecurity and uncertainty helps to instigate fear among in-group, which is arguably a relatively more effective strategy than instigating hatred against Tther group in the process of violence. This coalesces around narratives of blame which acts as 'fear speech' that instills fear among the in-group that the 'Other' poses an existential threat[71]. This becomes instrumental in mobilising the collective identity.

Narratives of blame refer to the Others' historically

[71]Buyse, A. (2014). Words of Violence: "Fear Speech," or How Violent Conflict Escalation Relates to Freedom of Expression. Human Rights Quarterly, 36(4), pp. 779 – 797. DOI: 10.1353/hrq.2014.0064.

subjugation of the in-group's nominal identity through oppression and persecution by the usurpation of power. Narrative and argumentation blends to highlight how this has contemporaneously been surreptitiously deployed to underline the legitimate claims of the in-group's towards territorial, social and cultural space.

This narrative of blame continues to align itself with contemporary events of public crises pivoting upon the Others who are portrayed as the harbingers of crises that endanger the safety of the general populace. Incidents of public crises are narrativized as a conspiracy by 'Others' and political entities contrary to the in-group's own. They are said to form a conspiratorial nexus to undermine the legitimate claims staked by the in-group.

Conspiratorial networks are invoked to highlight strategies that the Others are wily enough to leverage and forge to the detriment of the in-group. Popular media is often blamed for normalizing the representation of the Other while excluding the in-group's history and culture. This, the in-group argues, downplays the issues plaguing its wider communities and leadsing to a crisis in its way of life.

According to Ranajit Guha, narratives are sense-making organizational structures that have two functional component – one is indicative, the other is interpretive[72]. The indicative component serves the function of reportage or description while the interpretive component serves the function of explaining the above description[73]. These two components working together co-constitute meaning[74]. The first step entails the identification of two broad sequences, one of problem definition and the other, of response[75].

Thereafter each sequence so defined is constitutive of and bifurcates into a series of linked micro-sequences classified as per levels of analysis, from generalized (the narrative) to particularized (its component parts)[76]. The singular or combined effect of any number of such micro-sequences represents moments of context dependent risk or the potential to influence and alter linkages to subsequent micro-sequences[77].

Through narratives of blame, the in-group describes historical events and imbues them with explanatory value of subjugation and oppression of the in-group historically at the hand of the 'Other'. Subsequently, through techniques and strategies like co-opting contemporary public crises it builds sequences that lead into the overarching narrative of the need for renewed consciousness to right historical norms. Major contemporary events and isolated incidents are assimilated into the narrative as micro-sequences which work towards taking the narrative and its objective forward.

Internal equations

Use of the conceptual category of in-groups lends a sense of homegenisation to identity and practices. However, it is important to understand that within the workings of the in-groups competing variations exists in the form of online practices and offline presence. The in-group functions more like a network with nodes of influence which are sometimes loosely, sometimes more directly connected.

Facebook often becomes the platform for creating an active and interactive audience by these nodes. The various nodes within the in-groups have offline affiliations and presence of varying scales with efforts directed towards transforming online audiences into offline followers by invoking the need for direct action.

The nodes of the in-group network use different from of engagement – personal accounts, pages, and groups. Forms of engagement range from sharing details about their personal life apart from social and political ideologies. The nodes work towards various degrees of publicness.

Once personal account has enough number of followers, the next logical step becomes a page which has wider reach and easy access. Groups come into existence when they are created and provide a space for members to talk, post, voice their opinion, engage interactively, and have conversations.

However, the different nodes are not completely unknown to each other given that they often mention each other in either in cordiality or to make competing claims. Newer, often younger members, feel wider social media presence does not translate to direct action. Often sub-in-group identities create a friction within the in-group about who can stake a better claim to the purpose and objectives of the in-group.

Internal equation underscores that credibility does not rely on social media presence alone but how such narrativization is backed up with direct action. This leads to more prominent nodes having to negotiate the expectation their narrativization has engendered and the more real and culpable consequences of direct action.

[72]Guha, R. (1988). The Prose of Counter-Insurgency. In G. Chakravorty Spivak & R. Guha, Selected Sulaltern Studies. New York, Oxford: Oxford University Press.

[73]Ibid.

[74]Ibid

[75]Ibid.

[76]Ibid.

[77]Ibid

Networked leadership

While the section on internal equations attempted to explain the nature of associations – solidarities, competitiveness – that different online actors in these networks have with each other, networked leadership looks at how power is distributed within the spaces among the nodes occupying them. The premise for this section attempts to situate the observed speech acts and practices within the nodes disseminating it. This explains the nature of influence or leadership that arises out of decentralized networks.

Online spaces offer a way to rehearse social behaviour in a low-stakes environment. The power or symbolic-cultural capital that is being exerted through some of these groups and pages has the capacity to lead their audiences towards values espoused and propagated by the in-group. Repeated exposure to ideas by powerful individuals legitimize such ideas and have the capacity of translating calls to direct action to actual action.

In-group boundary making and collective identity building are inflected through emerging forms of leadership that is networked – speaking of not just associations between different kinds of actors but also how they are networked with the platforms they use, and with events of offline violence.

Networked leadership often leverages and mirrors the constellations of offline leadership with the potential to develop wider audiences for the in-group and articulate logics of belonging and action. However, the morphing of the social-cultural sphere into today's digitally mediated forms has meant that the landscape of this in-group assertion has mutated, as have the forms of hierarchy that it generates. In exploring these mutations, it has meant not only engaging with what the in-group's activism means as online linguistic expression, but also its cultural practice, in terms of the ways in which existing nodes devoted to the cause have adopted digital media technologies assimilating them within their offline work.

For most of these individuals and organizations, speech acts or calls for direct action exists as part of their wider project of performing the right or 'ideal' form of a member of the in-group, which involves several other duties such as urging other members to 'wake up' and organize, a worshipful reverence towards their accepted leadership, as well as the future of their territorial space, or even performing service towards their wider community. Their strategies towards audience engagement such as what they choose to highlight as what might be termed their specific brand of activism, is reflective of an internally consistent set of values that they are actively putting out through Facebook.

The influencers leading the conversation within the networked structure claim their affiliations to constellations of groups working on different scales and levels. They have high following and broadcast their services to the community, particularly during periods of public crises and distress. These efforts are livestreamed wherein it is reiterated how these are their 'natural duty' to be of service to the community by putting the interests of their communities above their own safety within a public crises. These regular broadcasts during a period of underpin the desire to prove offline, ground level carrying out of duty as a competing negotiation for attention.

In many instances, violence against the defined Other is valorised as service to the community. Calls to action and claiming of responsibility is a part of the leadership role that these nodes of influence have taken upon themselves. They operate both in an individual capacity disseminating content from their own profiles but also by running Facebook pages and groups as admin/moderators.

In the private groups studied, all had moderators who on their personal profiles claim affiliation to offline groups. It is important to note that moderators of private groups on Facebook have a more involved role than those of public groups, as they not only vet entry into the group but also are the first line of authority that is likely to deal with posts in the group being reported – unless the person reporting the content chooses to flag it straight to the platform instead of the moderators.

Apart from influencers negotiating and competing for an online audience and attention there are those with significant offline recognition with a digital presence that works to amplify the same. Live videos have often been the medium of choice through which calls to action have been made for the in-group to collectivise and arm themselves as service of the higher-order through which spatial and territorial claims of the in-group will be realised against social, cultural, discursive, and spatial incursion and usurpation of the Other.

This has often translated to direct calls to action that have given cause for extermination of the Other and how members of the in-group as non-institutional actors unlike politicians are better positioned to effect this change outside the purview of law. However, offline legal action against concerned members led to a wiping clean of the inciteful content such that it was not available to the public.

However, many of these platforms of engagement like pages and groups witness the sharing of content posted in-group conscience leaders identifiable by their honorifics who strongly advocate for a spatial-

territorial assertion that highlights the need for the in-group to arm and defend against the Other both internally and externally – i.e. within and outside spatial territorial boundaries. These are often individuals with large groups of followers offline and online – in one instance over 4 million and in a another instance a video which garnered more than 11,000 views within 24 hours as it remained on the platform.

However, scales of online activism involve different types of nodes which forge or amplify network and network connections. There are individuals, for example, who have gained prominence with nearly 0.7 million followers and whose live videos have boasted 100,000 in instances but who have been called out for not translating their online presence into offline direct action which acts as a proxy required to realise spatial-territorial claims.

Popularity on social media is not restricted to a single platform, a given node will have similar scales of popularity across other social media platforms. However, modes of engagement might be different across different platforms though Facebook lives continue to remain a preferred mode of direct engagement and interaction. Such popularity is also often buffeted by the followers constituting and maintaining online fan clubs and dedicated modes of engagement to engage and amplify with the rhetoric propounded by the principal node which also receives significant tractions in terms of engagement.

Popular nodes work towards dissemination of their views by substantiating them with argumentation, rhetorical strategies, and re-contextualised information. This is done through images, videos, text or through a multi-modal engagement which are often accompanied by direct calls to action, use of coded speech, and use of dehumanizing perjoratives. Group solidarity within the network come to the fore in instances of legal implications for the principal nodes which often involves hailing the individual so implicated as a defender of the in-group against the threats that it faces.

Scales of influence, rhetoric, and engaged support works to enable a position of power within networked leadership. This creates visibilities or a process of becoming which goes beyond being physically visible as a matter of gaining discursive attention and recognition[78]. Thus, visibility becomes something to be achieved like power, status, and authority[79]. Acquiring visibility also allows then to bestow visibility on certain aspects of the world thereby shaping discourses pertaining to them[80]. Visibility thus becomes more than the sensorial act of seeing and irreducible to the sensorium and carries with it the conditions that granted and maintains such visibility and increasingly attendant networked power that comes with it[81].

Instrumentalisation of virality

Virality in the context of digital media is usually invoked as a metaphor for scalability, in terms of the speed at which content spreads from one node to multiple others in the network as well as to multiplicity of social media platforms. It can be understood both as a phenomenon - a process taking place constantly in online spaces – as well as a tool for instrumentalization for strategies around. Taken together virality becomes more than an amorphous tendency within the network and assumes a form and direction.

Looking at virality within the wider information and content landscape like TV news cycles and Twitter trends allows us to make inferences about the level of normalization or validation of these ideas. This is because content or ideas that attain 'viral' status online are indicative of their collective modalities of visibility. When a particular type of call to action becomes viral with its associated content across multiple Facebook pages and groups, it is reflective of the readiness of these networks to justify it as acceptable or necessary.

"Popularity on social media is not restricted to a single platform, a given node will have similar scales of popularity across other social media platforms. However, modes of engagement might be different across different platforms though Facebook lives continue to remain a preferred mode of direct engagement and interaction."

[78]Chow, R. (2010). Postcolonial Visibilities: Questions Inspired by Deleuze's Method. In S. Bignall & P. Patton (Eds.), Deleuze and the Postcolonial, pp. 62 - 77. Edinburgh University Press: Edinburg.

[79]Ibid.

[80]Ibid.

[81]Ibid.

Viral content in association with an explicit call to action increases its potentiality to cause harm, not just through wider quantifiable acceptance but also the likelihood for it to get noticed and picked up as a talking point by public figures who can validate it through their social capital on different platforms.

Virality as scale depends on network effects with societal transformation from hierarchies to networks as the organizing principle of society[82];[83] . This leads to a more accommodative understanding of the increasingly mutually constitutive nature of the 'social' and the 'technological'. It also allows for a more granular understanding of how power is structured by, as well as within, this decentralized flow of information that enables the maintenance of networked society.

Virality is often instrumentalized for retributive justice where posts follow a formulaic progression that begins with a audio-visual content that depicts an instance where a given individual is seen saying or doing something that is offensive to the norms and beliefs of the in-group. This is often accompanied by captions that translate the phenomenon for the audiences which might also include a call for the post to 'be shared widely' or 'made viral' so that the offending individual is made a target of some form of retribution.

The more viral a post goes – including greater number of shares on groups with larger membership or by individual pages with a higher follower count – the more likelihood of the 'offender' being made to face these consequences. This has often resulted in online speech (in the comment sections as well as in captions by other people sharing the original post) translating to offline harassment or violence.

Virality is often instrumentalized for retributive justice where posts follow a formulaic progression that begins with a audio-visual content that depicts an instance where a given individual is seen saying or doing something that is offensive to the norms and beliefs of the in-group.

It was observed that the same content was often cross-posted across multiple groups. Given that many of these groups were observed to have common admins or moderators – the likelihood of affected direct action through a call increases – with the content reaching regional groups or groups in the same region or locale as the alleged offender. These are often followed by videos that show the alleged offender being brought to justice – oftentimes beaten or intimidated into apologising by a mob or a group of individuals who were able identify them or track them down.

Three characteristics of viral progression are noticeable in this context: (i) virality of the video containing content offensive to the in-group goes viral. This is at times accompanied by a call to action to make the content viral so that offender can be shamed, (ii) the second aspect takes places offline where the offender is identified and confronted by an angered group, (iii) is the return of this confrontation to the online space where the video of the confrontation goes viral where the alleged offender is seen apologizing and asking for forgiveness after either being beaten up or under duress.

At times the confrontation gets its digital life in an audio-visual format that shows the before and after, i.e. the alleged offense and then the retributive justice through confrontation. These help to reify the in-group pride, that rhetoric and narrative aims to build, are claimed through direct action. It further helps demonstrate how direct and violent actions are required to continually assert the claims made by the in-group against incursions as embodied by the alleged offensive actions.

Apart from direct violent action – confrontation can take place through legal implications against those who had committed the perceived offense as well as through targeted online campaigns which aim to intimidate the alleged offender into asking for forgiveness or back-tracking on their initial statements. Alleged offenders cut the socio-economic strata of the society. However, viral content is not only limited to individual instances of offense and violent action but also the need for institutional and legal reform in order to underscore the historical humiliation that the in-group has suffered. This helps the in-group articulate belonging within their spatial-territorial claims-making. The problematic with virality is that given the scale and velocity of the circulation of a given content it would have received enough traction even if was taken down and would have already been shared, screenshotted, downloaded – for all intents and purposes it has been immortalized, and can now be reproduced endlessly and edited multi-directionally.

[82]Castells, M. (2000). The rise of the network society. Blackwell Publishers: Oxford.

[83]Latour, B. (2005). Reassembling the social: An introduction to actor-network-theory. Oxford University Press: Oxford.

Posts that went viral and generated traction contains calls to action, misinformation, calls for economic boycott, and social exclusion/ segregation. Violent incidents widely shared on the platform and attributed to the Other were often made to trend without fact-checking and were picked by other networks and offline news media. However, when such instances came to be popularly debunked and internal fact-checking is forced to course correct – these instances do not go as viral as the corresponding misinformation campaign. This indicates that virality is undercut with the need to service the overarching narrative and to an extent given effect to intentionality.

However, virality is predicated not only on intentionality of networked actors but also technological materialities and algorithmic structures that incentivize such virality. Social media platforms operationalize the technologies that help facilitate the existence of networks. None of these exists in isolation of each other, yet are infrastructural to virality in differing ways. While networks form the conceptual lens with which to view social interaction, platforms and technologies form the material infrastructure of viral media and thus of viral hate speech.

Algorithms promote posts on the basis of greater user engagement and newsfeeds that are structured on opaque metrics of 'relevance', because of which there have been international reports on Facebook algorithms favoring extremist content[84]. Like Google's search algorithm or Netflix's recommendation algorithm, Facebook's newsfeed algorithm is a master algorithm made up of smaller sub-algorithms. Apart from a sorting algorithm, this also includes a complex relevancy algorithm that assigns a personalized relevancy score to every post that reaches a newsfeed, and then sorts the newsfeed on that basis. According to tech journalists and researchers, the number of variables – aspects of behavioural data being gathered – that go into calculating relevancy are in the hundreds[85].

Livestreaming online allows for individuals/ organizations to generate traction for their videos through directly demanding it from the audience. A popular strategy observed in the Facebook Live videos was that they often began with the individual actually spending the first few minutes repeatedly asking for the video to be 'made viral' and 'shared'. They appear to count on the loyalty of their viewers to ensure that they have shared it in as many like-minded groups as possible or their own individual profiles – and compliance is directly visible in the user engagement stats of such videos, as well as the validating responses in the comment sections.

Networked leaders have either significant follower counts of their individual pages who are willing promoters, or else tap into other networks of associations, such as members of the organization they represent; at times with several members regularly sharing each other's Live videos on various subjects, including multiple videos.

Another aspect of virality is the manufacturing and mostly planning the coordination of trending hashtags on Twitter, where hashtags are the primary drivers of virality, and harness the membership strength of the Facebook group to get things viral on Twitter[86]. The process is streamlined with different members dedicated to specific tasks, which not only includes composing content for a common pool of tweets, picking pictures, making memes but also monitoring the tweets that counter the hashtag they are amplifying that day, so that they can coordinate attacks against those Twitter users[87].

"Livestreaming online allows for individuals/organizations to generate traction for their videos through directly demanding it from the audience. A popular strategy observed in the Facebook Live videos was that they often began with the individual actually spending the first few minutes repeatedly asking for the video to be 'made viral' and 'shared'."

[84]Horowitz, J. & Seetharaman, D. (26 May 2020). Facebook executives shut down efforts to make the site less divisive. The Wall Street Journal. Retrieved from https://www.wsj.com/articles/facebook-knows-it-encourages-division-top-executives-nixed-solutions-11590507499 [24 October 2020].

[85]Oremus, W. (03 January 2016). Who controls your Facebook feed. Slate. Retrieved from http://www.slate.com/articles/technology/cover_story/2016/01/how_facebook_s_news_feed_algorithm_works.html?via=gdpr-consent [24 October 2020].

[86]Menghani, S.M. (27 May 2020). #OperationHashtag: How a Hindutva FB Group Pushes Politically Divisive Topics on Twitter. The Wire. Retrieved from https://thewire.in/politics/twitter-trends-manipulation-facebook-group-right-wing [24 October 2020].

[87]Ibid.

Virality can also take place within closed circuits of information such as a private Facebook group, where only members can see posts. This invisibly granted virality to certain posts is rather insidiously discerned by the code alone. When you open a Facebook group, the default setting in terms of filtering the posts that you are shown as a member, are the posts the algorithm has earmarked as 'Top Posts'. The other two options available for those that are aware of this policy and choose to switch from default, are 'Recent posts' ('see most recent posts first') and 'New activity' ('see posts with recent activity first'). Meanwhile, the description given for 'Top posts' states simply and without explanation: 'see the most relevant posts first'.

Lastly, in terms of deliberately maintaining virality, Facebook actually allows content creators that own Facebook pages to look at the 'viral reach' of their posts in order to track their performance analytics, so they can understand how to increase user engagement. By the social media company's definition, this is measured by the number of people who created a story from a post on your Facebook page, divided by the number of "unique people" who have seen that original post[88].

When you open a Facebook group, the default setting in terms of filtering the posts that you are shown as a member, are the posts the algorithm has earmarked as 'Top Posts'. The other two options available for those that are aware of this policy and choose to switch from default, are 'Recent posts' ('see most recent posts first') and 'New activity' ('see posts with recent activity first'). Meanwhile, the description given for 'Top posts' states simply and without explanation: 'see the most relevant posts first'.

Therefore, virality itself is a complex system structured through the platform's intelligent architecture and self/ collective intentionality. According to tech journalist Casey Newton, Facebook has responded to concerns raised about unchecked misinformation and disinformation about COVID-19 going viral on the platform, by saying they are working on a new approach that resembles a 'virality circuit breaker' - a suggestion given in a report on COVID-19 related misinformation by Center for American Progress.

The basic idea is to curb algorithmic amplification of trending COVID-19 virus content in order to streamline fact-checking of these posts. While this has been touted as a promising idea, there is no information yet on how effective Facebook's implementation of it will be, or whether there any plans to expand the idea to speech that is inciting violence or hate[89].

Virality as a repetitive symbolism is an idea based on wider understandings of virality in media, that look at it as the memetic repetition of desire[90] – which in the context of hate speech would include the continuous transmission of types of hateful imagery. There are a significant number of posts that are unlikely to get taken down due to their inexplicit nature of content, yet have potential to act as symbolic vectors of hate. Violence takes place through a large project of creating alternative versions of reality through explicit and implicit forms of messaging.

[88]Loomer, J. (25 September 2019). Hidden gems within Facebook Page Insights: Virality and viral reach. AgoraPulse. Retrieved from https://www.agorapulse.com/blog/facebook-page-insights-virality-viral-reach [24 October 2020].

[89]Newton, C. (20 August 2020). New ideas for fighting COVID-19 misinformation. The Interface. Retrieved from https://www.getrevue.co/profile/caseynewton/issues/new-ideas-for-fighting-covid-19-misinformation-272134 [24 October 2020].

[90]Parikka, J. (2007). Contagion and Repetition: On the Viral Logic of Network Culture. Ephemera, 7(2), pp. 287-308. Retrieved from http://www.ephemerajournal.org/sites/default/files/7-2parikka.pdf .

APPLICATION OF FRAMEWORK TO INTERNATIONAL INCIDENTS

The framework for social processes of direct action put forward an analytical framework built upon strategies instrumentalized towards certain calls to action through the uses and practices around content types that work towards maximizing engagement and reach.

This framework can be understood as the crystallization of the 'general' from a detailed archive of 'particulars' of ethnographic observation. In Collective Identity Building, in-group essentialism has been redefined as being analogous to spatial – territorial claims and exclusivist association leading to a systemic Otherisation through calls for institutional (legal reforms), structural (economic boycott), and violent exclusion (direct violent action).

Narratives of Blame addresses the large scale, multifaceted, scapegoating and stereotyping through the use of narrative and argumentation techniques to build and reify a discourse through different multiple strategies, and media platforms (both online and mainstream) and offline consequences of the same.

Networked Leadership and Internal Equations highlighted patterns of locations of power and the distribution of influence. The potentiality of speech to escalate to violence is directly tied to who is speaking as is also recognised by Facebook's policy of 'Dangerous Individuals and Organizations'[91].

Instrumentalizing Virality helps to visualize the ways in which these technologies contribute to generating and amplifying hate speech, including the scale of influence of online actors as well as the reach of specific instances of dangerous content.

With the nature of offline civic violence being predicated on group identity with the potential to cause large-scale civil unrest, the aim for this chapter is to understand whether these observations and patterns can be applied to other contemporary examples of violence escalating from online speech. Instead of drawing parallels with historical examples of hate speech that enabled mass violence such as Nazi Germany or the Rwandan genocide, it studies contemporary contexts of social media platforms being directly instrumentalized by perpetrators of offline violence.

This is done in order to contextualize the particularities of the increasingly networked relationship between reported implications of events of online speech escalating to offline violence all over the world. This chapter draws from popularly reported news sources and reports by international organisations to analyse given events and thereupon apply the framework to their specificities. It is by no means an exhaustive account of the said events. The purpose of this chapter is to perhaps test the application and validity of the framework in different contexts of offline civic violence and implication of social media within such violence, alongside their specific social histories and arrangement of power and social influence.

While Rwanda and Germany provide robust examples of hate speech patterns that have now been accepted as precursors to the eventual genocide, the proliferation of interactive media platforms such as Facebook created a radically different form of societal interaction and information flow – i.e. a radically different mode of cultural and ideological production. Mirroring this aspect of social media the civic violence engendered too is fragmented, networked, and atomised locally rather than a large wave of direct persecutory action by one group over another as in these two cases. This in a way makes it more temporally pervasive and normalised, thereby raising questions about moments of disjuncture wherein the societies are collectively able to move past narrativized and interpellated subjectivities.

This represents newer ways of conditions of civic violence to occur; mediated by social media platforms and augmented by technological materialities alongside discursive directionality provided the networked hierarchies of the in-group. Therefore, it becomes important to take into account the ways in which contemporary projects of concerted mass violence against minority groups have been enabled through social media platforms.

This chapter discusses the violence in Myanmar, Sri Lanka, and the lone wolf attack in Christchurch, New Zealand which brought the fore the role played by social media within social processes leading to offline violence and have led to international cognisance and even resolutions in the form of Christchurch call to eliminate terrorist and violent extremist content online.

In the case of Myanmar and Sri Lanka Facebook acknowledged the use of its platform to perpetrate

[91]Facebook Community Standards. (2020). Dangerous individuals and organisations. Retrieved from https://m.facebook.com/communitystandards/dangerous_individuals_organizations/ [18 September 2020]. Explained in greater detail in the following chapter on Facebook's existing Content Moderation policy.

targeted civic violence[92]. Sri Lanka's contemporary past saw the rise of radical groups, one of which – the Bodu Bala Sena (BBS) [trans. Buddhist Power Force] came under the scrutiny of the United Nations[93] which urged Sri Lanka to rein in 'faith-based violence'[94]. This was in the aftermath of a large protest rally staged by BBS in Aluthgama which resulted in inter-communal violence during which 4 people died and 80 were injured[95]. This was set in the context of growing attacks against minorities with 350 attacks against Muslims and 150 attacks against Christians reported in the two years preceding 2014[96].

Groups like the BBS have used made sustained use of social media – "posting memes, photos, videos and live broadcasts to spread and amplify their messages on a variety of platforms including Facebook, YouTube and Twitter" to mount speeches that have clear calls to actions for extermination and persecution of minorities with stereotyped epithets and misinformation.[97] In the 2018 communal violence that unfolded in Kandy, Facebook acknowledged that the proliferation of hate speech on its platforms may have contributed to the outbreak and escalation of the violence[98].

In Myanmar, the persecution and displacement of the Rohingya community coincided with democratisation of the country and a newly elected government. Democratisation of the country went hand in hand with the liberalisation of the telecommunications industry which led to the proliferating use of the internet and social media platforms[99]. This deep social media penetration has enabled charismatic leaders and groups to take advantage of social media to deepen communal fissures through divisive speech coupled with strategies of misinformation[100].

It has been accompanied by the rise in divisive rhetoric and communal violence post the beginning of reforms[101]. A Reuters investigation found "1,000 examples of posts, comments and pornographic images attacking the Rohingya and other Muslims on Facebook"[102]. The divisive messaging was both by organised groups as well as state officials[103]. According to report by the Médecins Sans Frontières (MSF) 6,700 Rohingya, including at least 730 children under the age of 5 were killed in the month after the violence broke out accompanied by reports of rape and sexual abuse and assault with 288 villages partially destroyed by fire[104].

> Groups like the BBS have used made sustained use of social media – "posting memes, photos, videos and live broadcasts to spread and amplify their messages on a variety of platforms including Facebook, YouTube and Twitter" to mount speeches that have clear calls to actions for extermination and persecution of minorities with stereotyped epithets and misinformation.

[92]Kamdar, B. (19 August 2020). Facebook's Problematic History in South Asia. The Diplomat. Retrieved from https://thediplomat.com/2020/08/facebooks-problematic-history-in-south-asia/ [18 September 2020].

[93]Kumar, S. (09 July 2014). The rise of Buddhist nationalism in Sri Lanka. The Diplomat. Retrieved from https://thediplomat.com/2014/07/the-rise-of-buddhist-nationalism-in-sri-lanka/ [12 October 2020].

[94]Srinivasan, N. (03 July 2014). U.N. urges Colombo to stop promotion of 'faith-based hatred'. The Hindu. Retrieved from https://www.thehindu.com/news/international/south-asia/un-urges-colombo-to-stop-promotion-of-faithbased-hatred/article6170959.ece [12 October 2020].

[95]Ibid.

[96]Ibid.

[97]Perera, A. & Rasheed, Z. (14 March 2018). Did Sri Lanka's Facebook ban help quell anti-Muslim violence. Al Jazeera. Retrieved from https://www.aljazeera.com/news/2018/3/14/did-sri-lankas-facebook-ban-help-quell-anti-muslim-violence [12 October 2020].

[98]Kamdar, B. (19 August 2020). Facebook's Problematic History in South Asia. The Diplomat. Retrieved from https://thediplomat.com/2020/08/facebooks-problematic-history-in-south-asia/ [18 September 2020].

[99]Office of the United Nations High Commissioner for Human Rights (OHCHR). (27 August 2018). Report of Independent International Fact-Finding Mission on Myanmar. Retrieved from https://www.ohchr.org/EN/HRBodies/HRC/MyanmarFFM/Pages/ReportoftheMyanmarFFM.aspx [11 September 2020].

[100]Ibid.

[101]Ibid.

[102]Reuters. (15 August 2018). Why Facebook is losing the war on hate speech in Myanmar. Reuters. Retrieved from https://www.reuters.com/investigates/special-report/myanmar-facebook-hate/ [12 October 2020].

[103]Office of the United Nations High Commissioner for Human Rights (OHCHR). (27 August 2018). Report of Independent International Fact-Finding Mission on Myanmar. Retrieved from https://www.ohchr.org/EN/HRBodies/HRC/MyanmarFFM/Pages/ReportoftheMyanmarFFM.aspx [11 September 2020].

[104]Ibid.

The lone wolf Christchurch attack by Brenton Tarrant helps to study this processual phenomenon of online speech and offline violence in the case of an individual actor in contrast to more coordinated and networked patterns of divisive speech employing multiple strategies of misinformation, dehumanisation, in-group glorification through overarching narrativization and argumentation. His use of Facebook live video feature to stream the attack online and a failure of not taking it down before the circulation of the attack video reached millions on Facebook and other social media platforms[105].

Tarrant also wrote a manifesto on Facebook which was a direct call to action[106]. Tarrant's act is important as it acted as inspiration for many other similar events[107];[108]. The manifesto filled with narrative of perceived fear of being wiped out by the minorities[109] is in line with other two larger examples of Myanmar and Sri Lanka. Tarrant's case provides a study of one person's perception built from the narrative of hatred and leading him to act alone. Unlike Myanmar and Sri Lanka, it may be a study of one single perpetrator but it is built on same idea of collective defence against an enemy as the other two largest country wide cases and hence becomes an important example to study on.

The inextricability of Facebook as a social media platform from its instrumentalisation as a vehicle for divisive rhetoric to grant social sanction of group-on-group civic and institutional violence across multiple locations requires us to test applicability of the framework of analysis derived from a specific site of observation to others where such phenomenon has taken place. This helps to bring to fore the applicability of the framework as a tool of referential analysis beyond its site of study.

Myanmar and the Rohingya crisis

In Myanmar, the persecution and displacement of the Rohingya community coincided with democratisation of the country and a newly elected government. However, the military retained administrative control with allocation of a quarter of all parliamentary seats[110]. This essentially made constitutional amendments impossible without the consent of the military[111]. Democratisation of the country went hand in hand with the liberalisation of the telecommunications industry which led to the proliferating use of the internet and social media platforms[112].

This development coupled with access to affordable devices and connectivity led to the wide-spread mobile and internet penetration and usage in Myanmar[113]. In such an environment, Facebook played a key role in structuring the experience of the internet for the country's population[114]. In partnership with Israeli start-up Snaptu, it made its platform available on basic feature phones which allowed it attain deeper penetration[115].

This deep social media penetration has enabled charismatic leaders and groups to take advantage of social media to deepen communal fissures through divisive speech coupled with strategies of misinformation[116]. It has been accompanied by the rise in divisive rhetoric and communal violence post

[105]Koh, Y. (2019). Why video of New Zealand massacre can't be stamped out. The Wall Street Journal. Retrieved from https://www.wsj.com/articles/why-video-of-new-zealand-massacre-cant-be-stamped-out-11552863615 [12 October 2020].

[106]Chung, A. (17 March 2019). New Zealand mosque shootings: Suspect's manifesto sent to PM's office minutes before attack. Sky News. Retrieved from https://news.sky.com/story/new-zealand-pm-to-discuss-live-streaming-with-facebook-11668059 [12 October 2020].

[107]Burke, J. (11 August 2019). Norway mosque attack suspect 'inspired by Christchurch and El Paso shootings'. The Guardian. Retrieved from https://www.theguardian.com/world/2019/aug/11/norway-mosque-attack-suspect-may-have-been-inspired-by-christchurch-and-el-paso-shootings [12 October 2020].

[108]O' Malley, N. (15 March 2020). Awaiting trial, the Christchurch attacker inspires a new global hatred. The Sydney Morning Herald. Retrieved from https://www.smh.com.au/world/oceania/awaiting-trial-the-christchurch-attacker-inspires-a-new-global-hatred-20200314-p54a2m.html [12 October 2020].

[109]Chung, A. (17 March 2019). New Zealand mosque shootings: Suspect's manifesto sent to PM's office minutes before attack. Sky News. Retrieved from https://news.sky.com/story/new-zealand-pm-to-discuss-live-streaming-with-facebook-11668059 [12 October 2020].

[110]Ebbinghausen, R. (2018). Myanmar's democracy movement 30 years on – military still calls the shots. DW. Retrieved from https://www.dw.com/en/myanmars-democracy-movement-30-years-on-military-still-calls-the-shots/a-44985212 [12 October 2020].

[111]Ibid.

[112]Office of the United Nations High Commissioner for Human Rights (OHCHR). (27 August 2018). Report of Independent International Fact-Finding Mission on Myanmar. Retrieved from https://www.ohchr.org/EN/HRBodies/HRC/MyanmarFFM/Pages/ReportoftheMyanmarFFM.aspx [11 September 2020].

[113]Paladino, B. (July 2018). Democracy disconnected: Social media's caustic influence on southeast Asia's fragile republics. Foreign Policy at Brookings. Retrieved from https://www.brookings.edu/wp-content/uploads/2018/07/FP_20180725_se_asia_social_media.pdf [12 October 2020].

[114]Ibid.

[115]Ibid.

[116]Ibid.

the beginning of reforms[117]. A Reuters investigation found "1,000 examples of posts, comments and pornographic images attacking the Rohingya and other Muslims on Facebook"[118]. The divisive messaging was both by organised groups as well as state officials[119].

In this context came the military crackdown in Rakhine State in 2017 to root out the ARSA (Arakan Rohingya Salvation Army) militants that led to the large-scale displacement of the Rohingya Muslim population[120]. The exodus came in the wake of an ARSA attack on 30 police posts which was followed by retaliation by the army[121]. According to witness testimonies, the retaliation included troops backed by local Buddhist mobs burning villages, attacking and killing civilians[122]. Prior to the 2017 crackdown there have reportedly been several waves of military action in the region[123].

According to report by the Médecins Sans Frontières (MSF) 6,700 Rohingya, including at least 730 children under the age of 5 were killed in the month after the violence broke out accompanied by reports of rape and sexual abuse and assault with 288 villages partially destroyed by fire[124]. The government of Myanmar denies citizenship to the Rohingyas who were also excluded from the 2014 census treating them as illegal immigrants from Bangladesh[125].

By end of 2018 over 7,25,000 people from the Rohingya community had fled to neighbouring Bangladesh out as a result of military operations[126]. In its report, the Office of the United Nations High Commissioner for Human Rights (OHCHR) said Myanmar army operations in the region involved carrying out of mass killings and gang rapes of Muslim Rohingya women with genocidal intent, further stating that the Commander-in-Chief and five generals should be prosecuted for orchestrating gravest crimes under law. [127]

In underscoring its inference of genocidal intent, the report pointed to the army Commander-in-Chief's statement that the "clearance operations" were not a response to a concrete threat from ARSA, but to the "unfinished job" of solving the "long-standing" "Bengali problem". However, a government appointed commission made selective admissions of wrong-doing by low-ranking army officials while summarily clearing the security forces of any massive violations[128].

The root cause of the conflict can be traced back to 1982 when a law cemented a stratified citizenship system that did not recognize Rohingyas as one of the 135 legally recognised ethnic groups of Myanmar[129]. Most Rohingya lack formal documents, and even those who come from families that have lived in Burma for generations do not have any way of providing "conclusive evidence" of their lineage in Burma prior to 1948, denying them Burmese citizenship[130]. Human Rights Watch, UN agencies, and others have long recognized the denial of citizenship as a root cause of the violence in Rakhine State.

[117]Ibid.

[118]Reuters. (15 August 2018). Why Facebook is losing the war on hate speech in Myanmar. Retrieved from https://www.reuters.com/investigates/special-report/myanmar-facebook-hate/ [12 October 2020].

[119]Office of the United Nations High Commissioner for Human Rights (OHCHR). (27 August 2018). Report of Independent International Fact-Finding Mission on Myanmar. Retrieved from https://www.ohchr.org/EN/HRBodies/HRC/MyanmarFFM/Pages/ReportoftheMyanmarFFM.aspx [11 September 2020].

[120]BBC Trending. (12 September 2018). The country where Facebook posts whipped up hate. Retrieved from https://www.bbc.co.uk/news/blogs-trending-45449938 [12 October 2020].

[121]BBC News. (23 January 2020). Myanmar Rohingya: What you need to know about the crisis. Retrieved from https://www.bbc.co.uk/news/world-asia-41566561 [12 October 2020].

[122]Ibid.

[123]Office of the United Nations High Commissioner for Human Rights (OHCHR). (27 August 2018). Report of Independent International Fact-Finding Mission on Myanmar. Retrieved from https://www.ohchr.org/EN/HRBodies/HRC/MyanmarFFM/Pages/ReportoftheMyanmarFFM.aspx [11 September 2020].

[124]BBC News. (23 January 2020). Myanmar Rohingya: What you need to know about the crisis. Retrieved from https://www.bbc.co.uk/news/world-asia-41566561 [12 October 2020].

[125]Ibid.

[126]Office of the United Nations High Commissioner for Human Rights (OHCHR). (27 August 2018). Report of Independent International Fact-Finding Mission on Myanmar. Retrieved from https://www.ohchr.org/EN/HRBodies/HRC/MyanmarFFM/Pages/ReportoftheMyanmarFFM.aspx [11 September 2020]

[127]Nebehay, S. (27 August 2018). Myanmar generals had "genocidal intent" against Rohingya, must face justice – UN. Reuters. Retrieved from https://www.reuters.com/article/myanmar-rohingya-un-idUSL8N1VH04R [11 September 2020]

[128]Human Rights Watch. (22 January 2020). Myanmar: Government Rohingya report falls short. Retrieved from https://www.hrw.org/news/2020/01/22/myanmar-government-rohingya-report-falls-short [15 October 2020].

[129]Office of the United Nations High Commissioner for Human Rights (OHCHR). (27 August 2018). Report of Independent International Fact-Finding Mission on Myanmar. Retrieved from https://www.ohchr.org/EN/HRBodies/HRC/MyanmarFFM/Pages/ReportoftheMyanmarFFM.aspx [11 September 2020].

[130]Human Rights Watch. (April 2013). "All You Can Do is Pray" Crimes Against Humanity and Ethnic Cleansing of Rohingya Muslims in Burma's Arakan State. Retrieved from https://www.hrw.org/reports/burma0413_FullForWeb.pdf [11 September 2020].

Role of Facebook

In Myanmar, Facebook is the internet – it is the most commonly used platform, often mobile phones comes installed the application[131]. A study on dangerous speech and offline violence in Myanmar attributes the widespread usage of Facebook to democratic reforms and liberalisation of the telecommunications industry, which is when smartphones and internet access actually became affordable for the average citizen[132]. In 2016, Facebook and Myanmar's largest operator Myanma Posts and Telecommunications (MPT) jointly launched "Free Basics" and "Facebook Flex"[133] . Soon it had become the primary mode of communication not just between citizens, but also between state and citizen, as Myanmar authorities used it as a tool with which they could regularly reach the public[134].

The OCHCR report mentions a "vast amount of hate speech across all types of platforms, including the print media, broadcasts, pamphlets, CD/DVDs, songs, webpages and social media accounts"[135]. It mentions having encountered, "over 150 online public social media accounts, pages and groups that have regularly spread messages amounting to hate speech against Muslims in general or Rohingya in particular".[136] According to OCHCR report,

Given Facebook's dominance in Myanmar, the Mission paid specific attention to a number of Facebook accounts that appear to be particularly influential considering the number of followers (all over 10,000, but some over 1 million), the high levels of engagement of the followers with the posts (commenting and sharing), and the frequency of new posts (often daily, if not hourly)[137].

Collective identity building and narratives of blame

The interactive or 'democratized' nature of social media platforms allows them to be studied as sites of identity construction – where construction can be understood as taking place discursively and symbolically. This process of constructing group identity typically manifests in the form of appealing to the collective self – i.e. that aspect of one's self-image that is derived from membership in social categories – in the groups and pages this study focused on, this was done to normalize the need for violence in service of the project of establishing and instituting spatial-territorial claims.

The nature of speech used for Rohingya Muslims is situated within the same nature of social exclusion that is inextricable from discriminatory articulations of belongingness and ownership in the context of land and territory. The specific strategies work towards discursively reifying collective identity through religious polarization, include dehumanizing language towards the Other, stereotyping, scapegoating the community for issues being faced by the country at large, misinformation and

"The interactive or 'democratized' nature of social media platforms allows them to be studied as sites of identity construction – where construction can be understood as taking place discursively and symbolically. This process of constructing group identity typically manifests in the form of appealing to the collective self – i.e. that aspect of one's self-image that is derived from membership in social categories – in the groups and pages this study focused on, this was done to normalize the need for violence in service of the project of establishing and instituting spatial-territorial claims."

[131]Stevenson, A. (06 November 2018). Facebook admits it was used to incite violence in Myanmar. The New York Times. Retrieved from https://www.nytimes.com/2018/11/06/technology/myanmar-facebook.html [15 October 2020].

[132]Fink, C. (17 September 2017). Dangerous Speech, Anti-muslim Violence, and Facebook in Myanmar. Journal of International Affairs. Retrieved from https://jia.sipa.columbia.edu/dangerous-speech-anti-muslim-violence-and-facebook-myanmar [11 September 2020].

[133]Office of the United Nations High Commissioner for Human Rights (OHCHR). (27 August 2018). Report of Independent International Fact-Finding Mission on Myanmar. Retrieved from https://www.ohchr.org/EN/HRBodies/HRC/MyanmarFFM/Pages/ReportoftheMyanmarFFM.aspx [11 September 2020].

[134]Ibid.

[135]Ibid.

[136]Ibid.

[137]Ibid.

disinformation that have the potential to incite action, and glorifying past incidents of violence against that community. These have been used to codify a sense of fear of 'the Other' through creating narratives of historical and contemporary injustice faced by the majoritarian forces.

A further examination of this core idea of a 'Muslim threat' endangering the Buddhist character of the country revealed certain underlying themes: presenting the Rohingya community as an existential threat to Myanmar, as a hindrance to racial purity of the country, and of Islam itself harming the sanctity of Buddhism as the dominant religion. The idea of an existential threat can be perceived through the messaging of Rohingyas as "illegal immigrants" that are "invading the country".

Phrases of Facebook posts studied by the OHCHR's Fact Finding Mission include "they will swallow us", "they sneak into the country", "boat people" (there is derogatory language attached to the term 'boat people' – the literal meaning refers to trash that floats along a river), "need to protect the Western Gate against a Muslim invasion", "they want to take away northern Rakhine as their independent state". Multiple reports mention perceptions of Rohingya people being inextricable from terrorism (for example, "Bengali extremist terrorists", "jihadists") and insecurity (for example, "criminals and rapists")[138] [139].

The described perception, one that has been validated through official Facebook pages of the government's channels of information, is that Myanmar's ethnic people should not tolerate mass illegal Muslim immigration, because "Bengali immigrants" or "terrorists" will violently alter the Buddhist character of the country and cause its demise – sometimes with references to Afghanistan or Indonesia, referring to them as countries were once Buddhist and are now majority Muslim[140]. Multiple instances of unsubstantiated Rohingya Muslim "terrorist plots" have also contributed to the narrative of the community being a threat.[141]

The perception of racial purity being threatened is furthered through messaging around population growth amongst the Rohingya as a problem for the country – Facebook posts described in multiple reports refer to "incontrollable birth rates", "they breed like rabbits", "extremely large families", the practice of polygamy and the negative consequences of inter-faith marriage[142]. Expressing religious hatred based on the idea that there is a community-wide conspiracy within Muslims to forcefully convert (women in particular) to Islam, as well as the stereotype of predatory Muslim men and physically abusive behaviour, is a theme that like all the other examples described in this section, is common.

Christina Fink's study of the role of Facebook in facilitating the spread of dangerous speech in the Rohingya genocide also mentions the spreading of false claims about high Muslim birth rates, increasing Muslim economic influence, and Muslim plans to take over the country, and forced inter-faith marriage and conversion of Buddhist women[143]. She describes how posts about these "increasing Muslim numbers" are often accompanied by gruesome images of ISIS brutality and selective photos from episodes of communal violence in Myanmar to suggest all Muslims are potential terrorists.

The idea of Myanmar's religious sanctity being threatened is described as being articulated through calls for institutionally curtailing Muslim traditions or customs, calling them incompatible with Buddhism. In fact, in 2015, the popular Buddhist outfit called the Organization for the Protection of Race and Religion (known as Ma Ba Tha), which has also been described as extremely active online, released a statement calling on the government to ban Muslims from slaughtering animals at religious events[144].

Further, there were calls for an economic boycott of Muslim-owned businesses and the promotion of the widespread use of stickers with Buddhist symbols to identify Buddhist-owned establishments.[145] Even

[138]Ibid.

[139]Human Rights Watch. (April 2013). "All You Can Do is Pray" Crimes Against Humanity and Ethnic Cleansing of Rohingya Muslims in Burma's Arakan State. Retrieved from https://www.hrw.org/reports/burma0413_FullForWeb.pdf [11 September 2020]

[140]Office of the United Nations High Commissioner for Human Rights (OHCHR). (27 August 2018). Report of Independent International Fact-Finding Mission on Myanmar. Retrieved from https://www.ohchr.org/EN/HRBodies/HRC/MyanmarFFM/Pages/ReportoftheMyanmarFFM.aspx [11 September 2020]

[141]Human Rights Watch. (April 2013). "All You Can Do is Pray" Crimes Against Humanity and Ethnic Cleansing of Rohingya Muslims in Burma's Arakan State. Retrieved from https://www.hrw.org/reports/burma0413_FullForWeb.pdf [11 September 2020]

[142]Office of the United Nations High Commissioner for Human Rights (OHCHR). (27 August 2018). Report of Independent International Fact-Finding Mission on Myanmar. Retrieved from https://www.ohchr.org/EN/HRBodies/HRC/MyanmarFFM/Pages/ReportoftheMyanmarFFM.aspx [11 September 2020]

[143]Fink, C. (17 September 2017). Dangerous Speech, Anti-muslim Violence, And Facebook In Myanmar. Journal of International Affairs. Retrieved from https://jia.sipa.columbia.edu/dangerous-speech-anti-muslim-violence-and-facebook-myanmar [11 September 2020]

[144]Office of the United Nations High Commissioner for Human Rights (OHCHR). (27 August 2018). Report of Independent International Fact-Finding Mission on Myanmar. Retrieved from https://www.ohchr.org/EN/HRBodies/HRC/MyanmarFFM/Pages/ReportoftheMyanmarFFM.aspx [11 September 2020]

[145]Fink, C. (17 September 2017). Dangerous Speech, Anti-muslim Violence, And Facebook In Myanmar. Journal of International

in the 2012 unrest and violence that set the stage for the events of 2017, two of the most influential groups in organizing anti-Rohingya activities i.e. the local order of Buddhist monks (the Sangha) and the locally powerful Rakhine Nationalities Development Party (RNDP – a political party formed by Arakanese nationalists), gave instructions to the Buddhist population to "socially and economically isolate the Rohingya", in order to cut off the remaining Muslims from basic services necessary for daily survival such as markets, food, and income-generating activities so that they would decide to leave[146].

Calls to boycott Muslim businesses were also amplified by the 969 movement (nationalist movement opposed to what they see as Islam's expansion in predominantly-Buddhist Myanmar [147]) and by the Ma Ba Tha.[148] State officials in Myanmar also advocated forms of violence or exclusion against Rohingya Muslims on Facebook, through personal profiles as well as public pages. The OHCHR report describes multiple instances of high-ranking officials equating the Rohingya population with terrorism in their Facebook posts.

Networked leadership

It is in recognition of the widespread usage of the platform amongst citizens, that apart from media outlets maintaining an online presence, the President, the State Counsellor, the Commander-in-Chief, the Ministry of Information, the army and other key governmental institutions also relied on Facebook to release news and information. The OHCHR report also pointed out that, "In a context of low digital and social media literacy, the Government's use of Facebook for official announcements and sharing of information[149] further contributed to users' perception of Facebook as a reliable source of information." The OHCHR report involved an in-depth examination of the communications issued by key government institutions in the months following the August 2017 events in Rakhine State. These include the official Facebook page of the Office of the Commander-in-Chief of all armed forces in Myanmar, Senior-General Min Aung Hlaing (2.9 million followers); the official State Counsellor's Information Committee Facebook page (almost 400,000 followers); the Senior-General Min Aung Hlaing's official Facebook page (1.4 million followers); and the official Ministry of Information Facebook page (1.3 million followers)[150].

This report attributes the nature, scale and organization of anti-Rohingya operations on the ground (the military crackdowns) to a level of preplanning and design on the part of the army leadership, which is consistent with the vision of state officials as seen from their Facebook posts. For example, the report mentions Commander-in-Chief, Senior-General Min Aung Hlaing, who stated at the height of the operations in September 2017, "The Bengali problem was a long-standing one which has become an unfinished job despite the efforts of the previous governments to solve it. The government in office is taking great care in solving the problem."[151] This post has since been taken down but the committee that wrote the report claims to have it on file.

Another example, dated 11 September 2017, the Office of the Commander-in-Chief's post states "Because they (the Rohingya) don't have citizenship, and they are not "Nationality" and not a recognized ethnic group, there is no way they could ask for a self-administered zone. That's why they will remove the governing structure (in Rakhine State) with whatever means possible. They will remove all the ethnic people, everyone except their own kind, in the region. They will make sure the government and other ethnic people cannot re-enter the region.". This was 'part 6' of a Facebook post on "Talk on Rakhine issue and security outlook" which is now removed, post described as being on file with the OHCHR Mission.[152]

Multiple reports mention state channels disseminating a constant stream of misinformation about events in Rakhine State, that downplayed the seriousness of the situation and misled domestic audiences. Allegations of serious human rights violations by the Myanmar security forces were systematically denied[153],

Affairs. Retrieved from https://jia.sipa.columbia.edu/dangerous-speech-anti-muslim-violence-and-facebook-myanmar [11 September 2020]

[146]Human Rights Watch. (April 2013). "All You Can Do is Pray" Crimes Against Humanity and Ethnic Cleansing of Rohingya Muslims in Burma's Arakan State. Retrieved from https://www.hrw.org/reports/burma0413_FullForWeb.pdf [11 September 2020]

[147]Marshall, A R C. (27 June 2013). Special Report: Myanmar gives official blessing to anti-Muslim monks. Reuters. Retrieved from https://www.reuters.com/article/us-myanmar-969-specialreport-idUSBRE95Q04720130627 [11 September 2020]

[148]Office of the United Nations High Commissioner for Human Rights (OHCHR). (27 August 2018). Report of Independent International Fact-Finding Mission on Myanmar. Retrieved from https://www.ohchr.org/EN/HRBodies/HRC/MyanmarFFM/Pages/ReportoftheMyanmarFFM.aspx [11 September 2020]

[149]Ibid.

[150]Ibid.

[151]Ibid.

[152]Ibid., p. 239.

[153]Office of the United Nations High Commissioner for Human Rights (OHCHR). (27 August 2018). Report of Independent International Fact-Finding Mission on Myanmar. Retrieved from https://www.ohchr.org/EN/HRBodies/HRC/MyanmarFFM/Pages/ReportoftheMyanmarFFM.aspx [11 September 2020].

while the Government spread "demonstrably false information about the situation in Rakhine State". The OHCHR report states that explicit calls for nationalist and patriotic action accompanied such narratives and misinformation, calls that suggested that the country is under siege, and at least implicitly encourages citizens to take action in their own hands.

Apart from the armed forces, research on the anti-Rohingya media narratives that enabled genocidal thinking, also focused on state media such as the newspaper Global New Light of Myanmar (GNLM) which maintained an active social media presence through posting articles daily and keeping online archives of front pages from the printed version[154] . GNLM articles not only often referred to Rohingya Muslims as "terrorists" or "militants", it also perpetuated fear of the community through emphasizing Muslim-majority areas as "camps of violent attackers"[155]. Through the Rohingya crisis in August-September 2017, 47 out of 53 front page articles by GNLM were focused on security concerns in Rakhine State, "overwhelmingly placing the blame for unrest on 'extremist terrorists'"[156].

The role of radical elements within the Buddhist monkhood in Myanmar in inciting violence against Rohingya Muslims has been greatly emphasized. The anti-Muslim '969 movement' created in 2012 and the Ma Ba Tha which emerged in 2014, have been two of the most active, well-resourced and effective in this regard. The report mentions,"High-profile monks, including Ashin Wirathu, Parmaukkha and Sitagu Sayadaw, have openly and actively espoused and promoted anti-Muslim narratives for many years."[157]

Ashin Wirathu, a Buddhist monk and a rigorous user of social media regularly made inflammatory poststo his more than 500,000 Facebook followers[158]. In 2014, Wirathu reposted on his Facebook page a report of a Buddhist female employee's rape by the Muslim proprietor of a local teashop in Mandalay[159]. Wirathu's post mentioned that he had called the proprietor to assert he would face justice, which was interpreted by some followers as a call to action. Crowds of agitated Buddhist men gathered in the streets of Mandalay, Muslims organized in defence, and fighting erupted. A Muslim man and a Buddhist man were killed. Later, the state media reported that the rape allegation was false.[160]

In 2017, a Muslim advisor to Aung San Suu Kyi was killed, and Wirathu praised her killers because their reasons were to do with 'extreme patriotism'. This led to the Buddhist monks' council banning him from giving public speeches for a year, but there was no curtailment of his online activities.[161] On his personal blog, Wirathu had also posted a series of videos entitled "Defend against the dangers of Jihad"; "Jihad and the future"; and "Jihad war and future Myanmar", all of which expressly called for action against the "immediate" Islamic threat facing the country.[162]

The fact that such as Ashin Wirathu were able to propagate hate speech on their Facebook pages while mass offline violence was underway, speaks to uneven application of laws by Facebook. In fact, Rohingya bloggers have claimed that Facebook has been quick to suspend or close their accounts for posting graphic photographs documenting the military's human-rights abuses and voicing criticism of the military. [163] It was reported Wirathu was permanently removed from Facebook in early 2018.[164]

Instrumentalisation of virality

The OHCHR and the Human Rights Watch reports suggested that, "outbreaks of violence have been preceded by visits or sermons of monks associated with the Ma Ba Tha, the distribution of anti-Muslim pamphlets and/or increased hate speech on social media."[165] One example of this, as mentioned briefly

[154]Lee, R. (2019). Extreme Speech in Myanmar: The Role of State Media in the Rohingya Forced Migration Crisis. International Journal of Communication. Retrieved from https://ijoc.org/index.php/ijoc/article/view/10123 [11 September 2020].

[155]Ibid.

[156]Ibid.

[157]Office of the United Nations High Commissioner for Human Rights (OHCHR). (27 August 2018). Report of Independent International Fact-Finding Mission on Myanmar. Retrieved from https://www.ohchr.org/EN/HRBodies/HRC/MyanmarFFM/Pages/ReportoftheMyanmarFFM.aspx [11 September 2020].

[158]Lee, R. (2019). Extreme speech in Myanmar: The role of state media in the Rohingya forced migration crisis. International Journal of Communication. 13. 3203-3224. Retrieved from file:///C:/Users/DEF/Downloads/10123-39500-1-PB.pdf. [11 September 2020].

[159]Fink, C. (17 September 2017). Dangerous speech, Anti-muslim Violence, and Facebook in Myanmar. Journal of International Affairs. Retrieved from https://jia.sipa.columbia.edu/dangerous-speech-anti-muslim-violence-and-facebook-myanmar [11 September 2020].

[160]Ibid.

[161]Ibid.

[162]Office of the United Nations High Commissioner for Human Rights (OHCHR). (27 August 2018). Report of Independent International Fact-Finding Mission on Myanmar. Retrieved from https://www.ohchr.org/EN/HRBodies/HRC/MyanmarFFM/Pages/ReportoftheMyanmarFFM.aspx [11 September 2020].

[163]Fink, C. (17 September 2017). Dangerous Speech, Anti-muslim Violence, And Facebook In Myanmar. Journal of International Affairs. Retrieved from https://jia.sipa.columbia.edu/dangerous-speech-anti-muslim-violence-and-facebook-myanmar [11 September 2020]

[164]Lee, R. (2019). Extreme Speech in Myanmar: The Role of State Media in the Rohingya Forced Migration Crisis. International Journal of Communication. Retrieved from

[165]Office of the United Nations High Commissioner for Human

above, is the violence that broke out in Mandalay in 2014 that resulted in two deaths, at least 20 people injured and significant property damage.[166]

It was reportedly triggered by Ashin Wirathu sharing an online news report from 30 June 2014, alleging that Muslim teashop owners had raped a Buddhist woman, identifying the teashop by name and even including its location and the full names of the alleged perpetrators and the victim. Wirathu's post on his Facebook page was captioned stating that the "Mafia flame (of the Muslims) is spreading" and that "all Burmans must be ready". Violence erupted the following day, but the rape story turned out to be false, with the "victim" reportedly admitting that she had fabricated the rape allegations. [167]

The role played by virality in continuous processes of dehumanization, stereotyping, divisive misinformation works through firstly, Facebook being a trusted and primary source of information for much of the public; and second, several important governmental and religious figures made use of the platform to push anti-Rohingya speech to their followers, as described earlier.

Anti-Muslim narratives in Myanmar comes together through a process of cross pollination of ideas between online and offline media (such as broadcast news agencies) with the overarching theme that Rohingya Muslims are illegal intruders that need to be dealt with through harsh state measures or violence, as well as the veneration of key leaders.

2018 civic violence in Sri Lanka

Following the end of the 33-year-old civil war between the State and The Liberation Tigers of Tamil Eelam (LTTE) in 2009 over the demand of the independent state for Tamils in Sri Lanka, the country has witnessed a gradual increase in violent attacks against religious sites and on religious communities, especially targeting Muslims[168] that form 9.7% of the country's total population. [169] This increase is visible in the form of growing numbers of demonstrations/ rallies, violent attacks and hate speech against the community[170].

In 2012, Buddhist monks destroyed a mosque in Dambulla, claiming it to be a violation of a Buddhist religious area[171]. In 2013, a clothes' warehouse owned by a Muslim businessman was targeted[172]. In the same year in August, the Masjid Deenul Islam (a Muslim prayer centre) in Sri Lanka's Grandpass was attacked during Maghrib (sunset prayers) by a mob of around 50-60 people, which later swelled to 200, reportedly led by Buddhist monks, leaving at least 5 people injured and several houses damaged ;[173]. [174] In order to contain the violence the police and the special task force imposed a curfew in the area[175]. The year 2014 also witnessed violence in the Aluthgama area of Sri Lanka in the form of riots that broke out after the Bodu Bala Sena (BBS) (founded in 2012) held a rally expressing anti-Muslim sentiments, leaving at least 4 dead and 80 injured and scores of Muslim owned homes and shops set ablaze, looted and destroyed[176].

Groups like BBS have enjoyed patronage of important political individuals[177]. As observed by the Secretariat for Muslims (SFM), a Muslim civil society organisation, there have been 538 anti-Muslim incidents recorded from 2013 to 2015[178]. Since 2012, the BBS had been involved in in direct action against the Muslims like raiding the Muslim-owned slaughter-houses claiming them to be breaking the law including demonstrating outside a law college, alleging exam results to be distorted in

Rights (OHCHR). (27 August 2018). Report of Independent International Fact-Finding Mission on Myanmar. Retrieved from https://www.ohchr.org/EN/HRBodies/HRC/MyanmarFFM/Pages/ReportoftheMyanmarFFM.aspx [11 September 2020].

[166]Mclaughlin, T. (7 June 2018). How Facebook's Rise Fueled Chaos and Confusion in Myanmar. Wired. Retrieved from https://www.wired.com/story/how-facebooks-rise-fueled-chaos-and-confusion-in-myanmar/ [12 September 2020].

[167]Office of the United Nations High Commissioner for Human Rights (OHCHR). (27 August 2018). Report of Independent International Fact-Finding Mission on Myanmar. Retrieved from https://www.ohchr.org/EN/HRBodies/HRC/MyanmarFFM/Pages/ReportoftheMyanmarFFM.aspx [11 September 2020].

[168] Aliff, S. M. (2015). Post-war conflict in Sri Lanka: Violence against Sri Lankan Muslims and Buddhist hegemony. International Letters of Social and Humanistic Sciences, 59, 109-125. Retrieved from doi:10.18052/www.scipress.com/ilshs.59.109 [13 October 2020].

[169]Ibid.

[170]Ibid.

[171]Subramanian, N. (07 March 2018). In Sri Lanka's anti-Muslim violence, an echo of post-war Sinhala triumphalism. The Indian Express. Retrieved from https://indianexpress.com/article/explained/sri-lanka-emergency-s-anti-muslim-violence-an-echo-of-post-war-sinhala-triumphalism-5088617/ [14 September 2020].

[172]Ibid.

[173]Farook, L. (31 August 2013). Calculated attack on Grandpass mosque. Colombo Telegraph. Retrieved from https://www.colombotelegraph.com/index.php/calculated-attack-on-grand-pass-mosque/ [14 September 2020].

[174]BBC. (11 August 2013). Sri Lanka Buddhist mob attacks Colombo mosque. Retrieved from https://www.bbc.com/news/world-asia-23653213 [13 October 2020].

[175]Ibid.

[176]Srinivasan, M. (11 July 2014). Attack in Aluthgama. Frontline. Retrieved from https://frontline.thehindu.com/world-affairs/attack-in-aluthgama/article6141587.ece [13 October 2020].

[177]Subramanian, N. (07 March 2018). In Sri Lanka's anti-Muslim violence, an echo of post-war Sinhala triumphalism. The Indian Express. Retrieved from https://indianexpress.com/article/explained/sri-lanka-emergency-s-anti-muslim-violence-an-echo-of-post-war-sinhala-triumphalism-5088617/ [14 September 2020].

[178]Ibid.

favour of the Muslims[179].

In the year 2018 two incidents of anti-Muslim violence were reported. First being in Ampara district on 26 February 2018 and second in Kandy district on 4-5 March 2018, leading to loss of lives and property of many[180]. Ampara district is a district with a near equal population of Muslims and Sinhala-Buddhists and is situated on the country's eastern coast[181]. These two particular incidents became an example of how social media was used to instrumentalise hate and offline violence, which compelled the authorities to block the access to social media and made Facebook to issue an apology for not taking timely action against the hateful content posted on their platform[182];[183].

Later in November 2019, ahead of the presidential elections in Sri Lanka, the official Facebook page of Gotabaya Rajapaksa (now: president of Sri Lanka) promoted a post propagating misinformation about "Muslim extremists" razing a Sri Lankan heritage site. And, even after AFP Sri Lanka's confirmation with the temple's chief monk that there had been no such attack, the post remained on Facebook[184].

While outlining suggestions and corrective measures in the anti-Muslim violence Brad Adams, the Asia Director of Human Rights Watch[185] said "Sri Lankan authorities need to do more than arrest those carrying out the anti-Muslim violence. They need to investigate and identify any instigators." He also established a link between the anti-Muslim violence and the role of the hardliner groups, "That means taking a hard look at the role and relationship between extremist Buddhist groups like the BBS and the Sri Lankan security forces[186]." While holding the Sri Lankan government responsible for the continuous attacks on the minorities he said, "The Rajapaksa government has long been ineffectual in holding those responsible for abuses to account[187]."

Facebook's Role and Apology

The Guardian's report[188] mentions about how a call for killing Muslims was allowed to remain on Facebook despite clearly violating the community guidelines: '"Kill all Muslims, do not spare even an infant, they are dogs," a Facebook status, white Sinhalese text against a fuchsia background, said on March 2018. This one-liner text post was a mix of both calling to take a violent action against Muslims by simply dehumanising them to be 'dogs'.

Appadurai argues how reducing the target communities to subhuman categories easily facilitates the work of large-scale murders as it creates a distance between the killers and the killed[189]. Report by Article One, a human rights consultancy, revealed that prior to the unrest in February and March 2018, Facebook had failed to take down hateful content that "resulted in hate speech and other forms of harassment remaining and even spreading" on the platform[190].

Facebook released an apology for not taking down the incendiary content and misinformation during the anti-Muslim violence in 2018 resulting to further provocation amongst people. "We deplore the misuse of our platform. We recognise, and apologise for, the very real human rights impacts that resulted", read the statement[191];[192] . They also said that in 2018 the mobs instrumentalized Facebook to coordinate attacks, and that the platform had "only two resource persons" to review content in Sinhala[193]. Similarly

[179]Haviland, C. (30 May 2015). The darker side of Buddhism. BBC. Retrieved, from https://www.bbc.com/news/magazine-32929855 [13 October 2020].

[180]Ibid.

[181]Ibid.

[182]Nazeer, T. (08 March 2018). Sri Lanka: Muslims fear more attacks during Friday prayers. Aljazeera. Retrieved, from https://www.aljazeera.com/news/2018/03/sri-lanka-muslims-fear-attacks-friday-prayers-180308155336083.html [14 September 2020].

[183]Aljazeera. (13 May 2020). Sri Lanka: Facebook apologises for role in 2018 anti-Muslim riots. Aljazeera. Retrieved, from https://www.aljazeera.com/news/2020/05/sri-lanka-facebook-apologises-role-2018-anti-muslim-riots-200513101243101.html [14 September 2020].

[184]Kamdar, B. (19 August 2020). Facebook's problematic history in South Asia. The Diplomat. Retrieved from https://thediplomat.com/2020/08/facebooks-problematic-history-in-south-asia/ [18 September 2020].

[185]Human Rights Watch. (19 June 2014).Sri Lanka: Justice key to end anti-Muslim violence. Retrived from https://www.hrw.org/news/2014/06/19/sri-lanka-justice-key-end-anti-muslim-violence [13 October 2020].

[186]Ibid.

[187]Ibid.

[188]Sayrah, A. D. (05 May 2018). Facebook helped foment anti-Muslim violence in Sri Lanka. What now?. The Guardian. Retrieved, from https://www.theguardian.com/commentisfree/2018/may/05/facebook-anti-muslim-violence-sri-lanka [14 September 2020].

[189]Appadurai, A. (2006). Fear of Small Numbers. In Fear of small numbers an essay on the geography of anger, pp. 49-85. Duke University Press:Durham.

[190]Aljazeera. (13 May 2020). Sri Lanka: Facebook apologises for role in 2018 anti-Muslim riots. Aljazeera. Retrieved, from https://www.aljazeera.com/news/2020/05/sri-lanka-facebook-apologises-role-2018-anti-muslim-riots-200513101243101.html [14 September 2020].

[191]Ibid.

[192]Nazeer, T. (15 May 2020). Facebook's Apology for its Role in Sri Lanka's Anti-Muslim Riots Should Spark Change. The Diplomat. Retrieved, from https://thediplomat.com/2020/05/facebooks-apology-for-its-role-in-sri-lankas-anti-muslim-riots-should-spark-change/ [14 September 2020].

[193]Aljazeera. (13 May 2020). Sri Lanka: Facebook apologises for role in 2018 anti-Muslim riots. Aljazeera. Retrieved, from https://www.aljazeera.com/news/2020/05/sri-lanka-facebook-apolo-

to Myanmar, the likelihood of harm being caused by divisive content online has been linked to Facebook's popularity as a platform in Sri Lanka, i.e. widespread usage due to factors such as affordable data packages allowing access to people across the socio-economic spectrum[194].

Collective identity and narratives of blame

A study conducted by the Centre for Policy Alternatives (CPA), Sri Lanka- Liking Violence: A study of Hate Speech of Facebook in Sri Lanka mentions about posts on Facebook directly targeting Muslims, revolve around 'revealing' the truth about Islam as being a religion that is intolerant towards 'polytheists' and oppressive towards women from their own community[195]. The content types of the posts are mainly curated in the form of infographics, stating Islam's full form as Intolerance, slaughtering, looting, arson and molestation of women[196]. The last bit extends further into separate infographics with Muslim burqa clad women crying on their fate of being born and married to a Muslim[197].

Such strategies to strengthen ideas based on the notions of the 'purity' of the group who need to have a homogenous approach towards their 'perceived enemies' also mirrors the beliefs of emboldening a collective identity that thrive within online spaces. The mobilization of feeling of 'we-ness' is a product of the idea of a sacred wholeness of the 'national demos' and the quantifiable or statistical idea of majority[198]. This is classified into the following three ways where a collective identity is created:

Fear mongering as a tactic: There lies a fear about Muslims taking over their countries by increasing their population. It is often strategized by describing one's community turning into a minority with only their current country having to live in and Muslims rapidly increasing their population like 'pigs'. This strategy of dehumanizing and reducing the targeted populations to subhuman category facilitates the large-scale mobilisation of calling to act against those communities.

According to the discourse created within the groups and pages, Muslims are procreating at a rate that they will no longer be in the minority and that one would be forced to adopt Islamic laws. In Sri Lanka's case there were open calls to take violent[199] action against Muslims by killing them. Anthropologist Arjun Appadurai[200] uses the term 'Predatory identities' for the majority community which fear turning into a minority and for this reason these predatory groups use pseudo-demographic arguments about rising birth rates of their targeted minority enemies. In Sri Lanka Muslims make for 9.7 percent[201] while Sinhala Buddhists make up 70.2 percent[202]. Therefore, such posts work to create a sense of urgency around a threat or what anthropologist Arjun Appadurai had called 'anxiety of incompleteness' which arises when there is a sensed lack of practical sovereignty[203].

Educative posts: In order to build and garner support from the members, content is presented as being long, elaborative and educational, professing to provide 'true' facts and figures. The posts are often shared with captions asking people to share it as much as possible for no mainstream educational system or media houses will show this 'real truth'. In Sri Lanka's case there have been elaborative posts revealing the 'truth' about specific incidents like the attack on the Grandpass Mosque in August 2013 as to how from the beginning, Muslims tricked the Buddhists and built an illegal mosque on the land of sacred Bo Tree and when it escalated to a religious conflict the mosque was declared legal.[204];[205] The content often caters to one-sided and disinformed point of view or news leading to the creation of a narrative of blame as the posts are often presented in conjunction with the ordeals faced by them.

gises-role-2018-anti-muslim-riots-200513101243101.html [14 September 2020].

[194]Sayrah, A. D. (05 May 2018). Facebook helped foment anti-Muslim violence in Sri Lanka. What now? The Guardian. Retrieved, from https://www.theguardian.com/commentisfree/2018/may/05/facebook-anti-muslim-violence-sri-lanka [14 September 2020].

[195]Ibid.

[196]Ibid.

[197]Ibid.

[198]Appadurai, A. (2006). Fear of Small Numbers. In Fear of small numbers an essay on the geography of anger, pp. 49-85. Duke University Press:Durham.

[199]Sayrah, A. D. (05 May 2018). Facebook helped foment anti-Muslim violence in Sri Lanka. What now? The Guardian. Retrieved, from https://www.theguardian.com/commentisfree/2018/may/05/facebook-anti-muslim-violence-sri-lanka [14 September 2020].

[200]Appadurai, A. (2006). Fear of Small Numbers. In Fear of small numbers an essay on the geography of anger, pp. 49-85. Duke University Press:Durham.

[201]US Department of State. (2019). 2018 Report on International Religious Freedom: Sri Lanka (Rep.). Retrieved from https://www.state.gov/reports/2018-report-on-international-religious-freedom/sri-lanka/ [14 September 2020].

[202]Ibid.

[203]Appadurai, A. (2006). Fear of Small Numbers. In Fear of small numbers an essay on the geography of anger, pp. 49-85. Duke University Press:Durham.

[204]Farook, L. (31 August 2013). Calculated Attack On Grandpass Mosque. Colombo Telegraph. Retrieved from https://www.colombotelegraph.com/index.php/calculated-attack-on-grand-pass-mosque/ [14 September 2020].

[205]Samaratunge, S., & Hattotuwa, S. (24 September 2014). Liking violence: A study of hate speech on Facebook in Sri Lanka (Rep.). The Centre for Policy Alternatives (CPA) Retrieved from https://www.cpalanka.org/liking-violence-a-study-of-hate-speech-on-facebook-in-sri-lanka/ [14 September 2020].

Group solidarity- There is a sense of urgency to protect the most vulnerable sections of one's own community- women and children- from the 'barbarism' of Muslims; reducing or dehumanising Muslim, especially men, to a sub-human category from whom the protection is needed. This gets strategized in the form of infographics or posters calling for actions like to 'wake up from your deep sleep', 'Wake up the Sinhalese'[206].'Sleep' is a constant metaphor that is used in both countries' cases as a wake-up call to the alleged atrocities inflicted over one's community at the hands of Muslims.

Networked Leadership

The tensions between the two communities heightened in 2012 with the rise of radical groups like BBS[207]. This group used fear tactics about Muslims taking over the country through their dominance in economic and demographic spheres[208] and even attempted to abolish Halal certification[209] for food and other products manufactured in Sri Lanka[210] in July 2013. These calls for economic boycott are reflective of a distinct pattern of calls for action that were also identified during ethnographic observation.

CPA's study[211] carried out a content analysis of over 20 extreme Buddhist Facebook pages in the backdrop of increasing incidents of hate crimes against Muslims starting from August 2013 attack on Masjid Deen-ul-Islam in Grandpass Colombo and then in June 2014 attack on Muslims in Aluthgama in South Sri Lanka.

The study also mentions Sinhala-Buddhist group like BBS and their anti-Muslim rhetoric in both online (social media) and offline spaces. The usage of Facebook to disseminate hate against the minority has been termed as "hate, hurt and harm"[212]. Sanjana Hattotuwa, an analyst at the Centre for Policy Alternatives (CPA) in Sri Lanka's Colombo, said that since the Sinhala Buddhist nationalist individuals and groups are "technologically savvy", they have been using social media to spread and amplify their messages against Muslims on social media platforms like Facebook including the Facebook pages belonging to Amith Weeransighe and also groups such as BBS and Sinhala Ravaya[213].

Weeransighe is one of the prominent figures who was arrested over the anti-Muslim violence which occurred in 2018. In a video posted shortly before the riots he was spotted urging his followers to gather towards the Digana area in the Kandy district: "This town has come to belong only to the Muslims. We should have started to address this a long time ago[214]".

The CPA report establishes that the violence that broke out in June 2014 in Aluthgama was a direct result of BBS' General Secretary Ven. Galagoda Aththe Gnanasara's speech that was delivered in a public rally prior to the violence where he had uttered that 'if any Sinhalese gets touched by a 'Marakkalaya' (trans. Muslims) that will be the end of everyone'[215;216].

In 2014, BBS invited Myanmar's hardliner Buddhist monk Ashin Wirathu to a rally in Colombo ahead of the presidential election, where Wirathu said he would join hands with the BBS to "protect" Buddhists[217]. BBS supporters often target the Buddhist priests who support religious harmony and criticise the BBS. As observed by the report[218] the

[206]Ibid.

[207]Gunasingham, A. (2018). Arrest of Influential Religious Hardliner and Religious Extremism in Sri Lanka. Counter Terrorist Trends and Analyses, 10(8), 7-9. Jstor. Retrieved, from https://www.jstor.org/stable/26481828 [14 September 2020].

[208]Ibid

[209]Wong, J., & Millie, J. (15 February 2015). Explainer: What is halal, and how does certification work? The Conversation. Retrieved, from https://theconversation.com/explainer-what-is-halal-and-how-does-certification-work-36300 [14 September 2020]

[210]Gunasingham, A. (2019). Buddhist Extremism in Sri Lanka and Myanmar: An Examination. Counter Terrorist Trends and Analyses, 11(3), 1-6. Jstor. Retrieved, from https://www.jstor.org/stable/26617827 [14 September 2020].

[211]Samaratunge, S., & Hattotuwa, S. (24 September 2014). Liking violence: A study of hate speech on Facebook in Sri Lanka (Rep.). The Centre for Policy Alternatives (CPA) Retrieved from https://www.cpalanka.org/liking-violence-a-study-of-hate-speech-on-facebook-in-sri-lanka/ [14 September 2020].

[212]Executive Director of the Centre for Policy Alternatives Dr. Paikiasothy Saravanamuttu calls this phenomenon of disseminating hate speech through a range of media around the growth of Islamophobia in post-war Sri Lanka as 'hate, hurt and harm'.

[213]Perera, A. & Rasheed, Z. (14 March 2018). Did Sri Lanka's Facebook ban help quell anti-Muslim violence. Al Jazeera. Retrieved from https://www.aljazeera.com/news/2018/3/14/did-sri-lankas-facebook-ban-help-quell-anti-muslim-violence [13 October 2020].

[214]Ibid

[215]Samaratunge, S., & Hattotuwa, S. (24 September 2014). Liking violence: A study of hate speech on Facebook in Sri Lanka (Rep.). The Centre for Policy Alternatives (CPA) Retrieved from https://www.cpalanka.org/liking-violence-a-study-of-hate-speech-on-facebook-in-sri-lanka/ [14 September

[216]Pieris, S. (06 July 2014). Hate speech — sowing the dragon's teeth. The Sunday Times. Retrieved from http://www.sundaytimes.lk/140706/sunday-times-2/hate-speech-sowing-the-dragons-teeth-105806.html [13 October 2020].

[217]Subramanian, N. (07 March 2018). In Sri Lanka's anti-Muslim violence, an echo of post-war Sinhala triumphalism. The Indian Express. Retrieved from https://indianexpress.com/article/explained/sri-lanka-emergency-s-anti-muslim-violence-an-echo-of-post-war-sinhala-triumphalism-5088617/ [14 September 2020]

[218]Samaratunge, S., & Hattotuwa, S. (24 September 2014). Liking violence: A study of hate speech on Facebook in Sri Lanka (Rep.). The Centre for Policy Alternatives (CPA) Retrieved from https://

critics of the BBS are often called out as 'traitors' who feed on 'Muslims' money' and hence talk in favour of Muslims, betraying Buddhism. Such posts are curated in the form of infographics with long elaborative captions explaining how such people are whitewashing the 'extremism' of Muslims and are rather cowards who cannot stand for their own religion.

Instrumentalising Virality

Towards the end of February 2018 anti-Muslim violence broke out in two different parts of Sri Lanka, in Ampara district on 26 February 2018 and Kandy district on 4-5 March 2018[219]. This led to vandalisation of two mosques, shops and other buildings in Kandy, and two mosques and shops in Ampara at the hands of Sinhalese Buddhist mobs. The violence took the lives of 2 people and left 5 injured. In response, the government imposed emergency and curfew to contain the violence and suspended internet services in the affected areas and even blocked access[220] to Facebook in an attempt to halt the organisers from planning more violence and spreading false rumours[221].

The violence in Ampara district was caused due to a misinformation campaign around a Muslim restaurant owner for allegedly mixing sterilisation drugs in the food, this was later proven false at two levels- first, what was assumed to be sterilization pill, was merely a small ball of dough; second, the doctors stepped in and clarified that scientifically there is no pill that can render its user permanently sterile[222].

On 26 February 2018, a mob shot a video of forceful confession of the restaurant owner- A.L.Farsith- for mixing 'wandapethi' (sterilisation drugs) in the food he served at his restaurant[223]. Farsith said that his Sinhalese is not very good and hence, he nodded his head out of fear when asked about the accusation of mixing pills in food[224]. The virality of the video was such that it led to calls for taking violent action against the shop owner[225]. In this incident not only his shop was vandalised by the mob but the video went viral on social media, aggravating to anti-Muslim violence that lead to mosques and vehicles being gutted[226].

The violence in Kandy on 4-5 March was a result of a traffic accident that happened between 4 Muslim men and a Buddhist truck driver (H.G Kumarasinghe) on 22 Feb 2018[227]. The clash further led to the death of the truck driver, this incident later spiralled down to calls for retribution and anti-Islam polemics flooding social media and eventually taking the form of riot 11 days after the death of Kumarasinghe[228] i.e. on 4-5 March 2018[229]. Then President Maithiripala Sirisena, in an interview with Sinhala weekly Divaina, blamed social media for the riots: "Extremist groups were using social media in the most heinous manner, That is why we had to limit it[230]."

According to the CPA report[231], posts are often in Sinhalese language which slips Facebook's radar of content moderation of . Another report by CPA on Confronting Accountability for Hate Speech in Sri Lanka: A Critique of the Legal Framework[232] shows the inability and unwillingness of the authorities to prosecute perpetrators of hate speech under the existing, 'International Covenant on Civil and Political Rights Act 56 of 2007' that falls under Sri Lanka's existing legal framework addressing Hate speech.[233]

www.cpalanka.org/liking-violence-a-study-of-hate-speech-on-face-book-in-sri-lanka/ [14 September 2020]

[219]Subramanian, N. (07 March 2018). In Sri Lanka's anti-Muslim violence, an echo of post-war Sinhala triumphalism. The Indian Express. Retrieved from https://indianexpress.com/article/explained/sri-lanka-emergency-s-anti-muslim-violence-an-echo-of-post-war-sinhala-triumphalism-5088617/ [14 September 2020]

[220]Nazeer, T. (08 March 2018). Sri Lanka: Muslims fear more attacks during Friday prayers. Aljazeera. Retrieved, from https://www.aljazeera.com/news/2018/03/sri-lanka-muslims-fear-attacks-friday-prayers-180308155336083.html [14 September 2020].

[221]Ibid.

[222]Borham, M., & Attanayake, D. (03 March 2018). Tension in Ampara after fake 'sterilization pills' controversy. Sunday Observer. Retrieved, from http://www.sundayobserver.lk/2018/03/04/news/tension-ampara-after-fake-%E2%80%98sterilization-pills%E2%80%99-controversy [14 September 2020].

[223]Ibid.

[224]Ibid.

[225]Ibid.

[226]Ibid.

[227]Allard, T., & Aneez, S. (25 March 2018). Police, politicians accused of joining Sri Lanka's anti-Muslim riots. Reuters. Retrieved, from https://www.reuters.com/article/us-sri-lanka-clashes-insight/police-politicians-accused-of-joining-sri-lankas-anti-muslim-riots-idUSKBN1H102Q?il=0 [14 September 2020].

[228]Ibid.

[229]Subramanian, N. (07 March 2018). In Sri Lanka's anti-Muslim violence, an echo of post-war Sinhala triumphalism. The Indian Express. Retrieved from https://indianexpress.com/article/explained/sri-lanka-emergency-s-anti-muslim-violence-an-echo-of-post-war-sinhala-triumphalism-5088617/ [14 September 2020].

[230]Perera, A. & Rasheed, Z. (14 March 2018). Did Sri Lanka's Facebook ban help quell anti-Muslim violence. Al Jazeera. Retrieved from https://www.aljazeera.com/news/2018/3/14/did-sri-lankas-facebook-ban-help-quell-anti-muslim-violence [12 October 2020].

[231]Samaratunge, S., & Hattotuwa, S. (24 September 2014). Liking violence: A study of hate speech on Facebook in Sri Lanka (Rep.). The Centre for Policy Alternatives (CPA) Retrieved from https://www.cpalanka.org/liking-violence-a-study-of-hate-speech-on-face-book-in-sri-lanka/ [14 September 2020].

[232]Centre For Policy Alternatives. (September 2018). Confronting Accountability for Hate Speech in Sri Lanka: A Critique of the Legal Framework. Retrieved from https://www.cpalanka.org/confronting-accountability-for-hate-speech-in-sri-lanka-a-critique-of-the-legal-framework/ [12 October 2020].

[233]Ibid.

> Another report by CPA on Confronting Accountability for Hate Speech in Sri Lanka: A Critique of the Legal Framework shows the inability and unwillingness of the authorities to prosecute perpetrators of hate speech under the existing, 'International Covenant on Civil and Political Rights Act 56 of 2007' that falls under Sri Lanka's existing legal framework addressing Hate speech.

Lone Wolf Attacks: Christchurch, New Zealand

The 'lone wolf' is a commonly used term referring to a violent act committed by a single perpetrator motivated by ideological reasons. The term defines the act planned and executed by a single person without the external support of other individuals, organization or the government[234]. While, the attacks are conducted independently, but there may be possibility of a connection in the form of source of radicalization and instruction through a formal network in form of an organized group.[235] Tarrant was connected to a number of Austrian far right groups who had invited him to visit and had received donations from him.[236]

The term lone wolf has existed in America at least since 1940s but it was popularized by white supremacists Tom Metzger and Alex Curtis in the 1990s[237] using the then new technology of internet through his online magazine Nationalist Observer.[238] In fact, the Federal Bureau of Investigation's joint investigation with San Diego Police against Alex Curtis was named 'Operation Lone Wolf' because of Curtis' encouragement to the idea of lone wolf activism in support of achieving the goal by any means necessary.[239]

Lone wolf attacks have seen a rise in the twenty-first century[240]. This period coincides with the rise of technology and internet access around the world. The democratised nature of the internet as a source of information, which at times can enable lone wolves to prepare terrorist attacks and act truly independently in terms of acquiring resources online, learning to use and practicing the weapons, as well as planning and executing the attack[241]. Online social media platforms have become an avenue for attracting potential members and followers for the organization.[242]

On 15 March 2019, a 'white supremacist' Brenton Tarrant carried out two mass shootings at two different mosques in Christchurch, New Zealand during a Friday prayer leaving 51 people dead and many other injured.[243] The attack took place at the time of Friday prayers and went on for 18 minutes from the first call to police till the arrest of Tarrant. The first 17 minutes of the attack were streamed live on Facebook through its live video feature and recorded through a head cam by Tarrant.[244] Tarrant shared an 87-page manifesto consisting of his ideological, political leanings along with

[234]Hamm, M. (2013). Lone wolf terrorism in America: Using knowledge of radicalization pathways to forge prevention strategies. National Institute of Justice. YouTube video. (:26-1:00 mins.). Retrieved from https://www.youtube.com/watch?v=px-lhuA1ZgA. [15 September 2020].

[235]Capellan, JA. (2018). Killing alone: Can the work performance literature help us solve the enigma of lone wolf terrorism?, in Terrorism in America, ed. Robin Maria Valeri and Kevin Borgeson. New York: Routledge. Retrieved from https://www.taylorfrancis.com/books/e/9781315456010/chapters/10.4324/9781315456010-9. [15 September 2020].

[236]Wilson, J. (16 May 2019). Christchurch shooter's links to Austrian far right 'more extensive than thought'. The Guardian. Retrieved from https://www.theguardian.com/world/2019/may/16/christchurch-shooters-links-to-austrian-far-right-more-extensive-than-thought. [16 September 2020].

[237]Weimann, G., 2012. Lone wolves in cyberspace. Journal of Terrorism Research, 3(2). Retrieved from http://doi.org/10.15664/jtr.405. [15 September 2020].

[238]ADL. (2000). Alex Curtis: 'Lone wolf' of hate prowls the Internet. Retrieved from https://www.adl.org/sites/default/files/documents/assets/pdf/combating-hate/Alex-Curtis-Report.pdf. [15 September 2020].

[239]FBI. Operation lone wolf. Retrieved from https://archives.fbi.gov/archives/sandiego/about-us/history/operation-lone-wolf. [15 September 2020].

[240]Worth, K. (14 July 2016). Lone wolf attacks are becoming more common — and more deadly. PBS. Retrieved from https://www.pbs.org/wgbh/frontline/article/lone-wolf-attacks-are-becoming-more-common-and-more-deadly/. [15 September 2020].

[241]Benson, D.C.(2014). Why the Internet is not increasing terrorism. Security Studies 23: 293–328. Retrieved from https://doi.org/10.1080/09636412.2014.905353. [15 September 2020].

[242]Weimann, G. (2010). "Terrorist Facebook: terrorist and online social networking." In: Web intelligence and security (Eds., M. Last and A. Kandel). Amsterdam: NATO Science for Peace and Security Series, pp. 19-30. Retrieved from file:///C:/Users/DEF/Downloads/405-951-1-PB%20(1).html. [15 September 2020].

[243]BBC. (15 March 2019). Christchurch shootings: 49 dead in New Zealand mosque attacks. Retrieved from https://www.bbc.com/news/world-asia-47578798. [15 September 2020].

[244]Wakefield, J. (16 March 2019). Christchurch shootings: Social media races to stop attack footage. BBC. Retrieved from https://www.bbc.com/news/technology-47583393. [15 September 2020].

popular culture references.[245] During the live video of shooting, he shared popular culture memes and played xenophobic Serbian song called 'Remove Kebab'.[246]

Tarrant's idea behind the attack in New Zealand, as he explained, in his manifesto was "to create conflict between the two ideologies within the United States on the ownership of firearms. He believed that he represented the "millions of white men who created America" and thought that through his action, the gun law debate would further the social, cultural, political and racial divide." Thus "ensuring a civil war in America which he called the death of the 'melting pot' pipe dream"[247]. He said his inspiration was Anders Behring Breivik who was convicted of a similar mass attack in Norway in 2011.[248] Tarrant referred to him as Knight Breivik in his manifesto. [249]

Tarrant's use of social media especially a messaging board called 8chan where he posted his intentions in an 87 pages long manifesto[250] and his live streaming of the attacks indicate social media's connection in the violence. Tarrant said in his manifesto that he had received, developed and researched his beliefs on the internet and that was the only place to find the 'truth'[251]. Tarrant's Facebook account had been taken down after the attack.

First to focus on the content type section of the research, there are only two posts that are in public knowledge with which comparison can be drawn. Firstly, the manifesto posted before the attack carried the nature of a long post, albeit it is considerably longer than any standard post on Facebook. Second was his live video of the attack that shook the whole world. The feature of live video used by Tarrant to live stream the attack shows the misuse of the feature.

As witnessed in ethnography, live video cannot be taken down unless they are reported widely,[252] Tarrant managed to stream first 17 minutes of his attack which fully covered the first attack at Al-Noor Mosque. Facebook took the video down after police reported it to them.[253] Even after blocking his profile, the video was circulating through various social media accounts. Reportedly 1.5 million digital copies were taken down by Facebook in the first 24 hours.[254] New Zealand's Prime Minister Jacinda Ardern was asked whether the live video feature should be stopped by Facebook.[255] In the aftermath of Tarrant's attack, Australian parliament passed a new social media law to penalize volatile content.[256]

"Firstly, the manifesto posted before the attack carried the nature of a long post, albeit it is considerably longer than any standard post on Facebook. Second was his live video of the attack that shook the whole world. The feature of live video used by Tarrant to live stream the attack shows the misuse of the feature."

[245]Kirkpatrick, D. (15 March 2019). Massacre suspect traveled the world but lived on the Internet. The New York Times. Retrieved from https://www.nytimes.com/2019/03/15/world/asia/new-zealand-shooting-brenton-tarrant.html. [15 September 2020].

[246]Schindler, J.R. (20 March 2019). Ghosts of the Balkan wars are returning in unlikely places. The Spectator. Retrieved from https://spectator.us/ghosts-balkan-wars-returning/. [15 September 2020].

[247]Kirkpatrick, D. (15 March 2019). Massacre suspect traveled the world but lived on the Internet. The New York Times. Retrieved from https://www.nytimes.com/2019/03/15/world/asia/new-zealand-shooting-brenton-tarrant.html. [15 September 2020].

[248]McIntyre, J. (17 January 2012). Anders Behring Breivik: a disturbing ideology. The Independent. London. Retrieved from http://blogs.independent.co.uk/2011/07/25/anders-behring-breivik-a-disturbing-ideology/. [16 September 2020].

[249]Tarrant, B. (2019) The Great Replacement. Retrieved from https://www.docdroid.net/48JypPr/the-great-replacement-original-by-brenton-tarrant-pdf. [14 September 2020].

[250]Wong, J.C. (5 August 2019). 8chan: the far-right website linked to the rise in hate crimes. The Guardian. Retrieved from https://www.theguardian.com/technology/2019/aug/04/mass-shootings-el-paso-texas-dayton-ohio-8chan-far-right-website. [16 September 2020].

[251]Tarrant, B. (2019) The Great Replacement. Retrieved from https://www.docdroid.net/48JypPr/the-great-replacement-original-by-brenton-tarrant-pdf. [14 September 2020].

[252]Doffman, Z. (24 March 2019). Facebook admits it can't control Facebook Live - Is this the end for live streaming. Forbes. Retrieved from https://www.forbes.com/sites/zakdoffman/2019/03/24/could-this-really-be-the-beginning-of-the-end-for-facebook-live/#-5070879fac8b. [18 September 2020].

[253]Chung, A. (17 March 2019). New Zealand mosque shootings: Suspect's manifesto sent to PM's office minutes before attack. Sky News. Retrieved from https://news.sky.com/story/new-zealand-pm-to-discuss-live-streaming-with-facebook-11668059. [16 September 2020].

[254]Ibid

[255]Ibid

[256]Karp, P. (4 April 2019). Australia passes social media law penalising platforms for violent content. The Guardian. Retrieved from https://www.theguardian.com/media/2019/apr/04/australia-passes-social-media-law-penalising-platforms-for-violent-content . [15 September 2020]

Whereas, the New Zealand government in collaboration with the French government founded Christchurch Call two months after the attack on 15th May 2019[257]. This was founded initially with the support of governments of 17 countries increasing to support of 31 more countries and international organization such as UNESCO and Council of Europe in September 2019, along withonline service providers such as Google, Amazon and Facebook. According to the call the governments are to ensure effective enforcement of the laws to counter terrorism and violent extremism and online service providers pledged to provide more transparency in the setting of community standards or terms of services.[258]

Brenton Tarrant's manifesto and live video, both had a heavy use of memes and references to memes. The weaponry shown in live video had racist and xenophobic memes pasted in the form of stickers and hate speech written over them with a white marker.[259] . His manifesto carried news links to selected instances of crimes against European Christians. He also posted links to videos from some Facebook account in his manifesto accusing Muslims of grabbing Christian land.[260] He said in his manifesto, that those videos proved that Muslims were aware of their guilt.[261]

Tarrant's manifesto relied heavily on disinformation as a strategy. The manifesto repeatedly argues about a perceived high fertility rate of Muslims which will lead to replacement of white Europeans. Tarrant declined to call himself an Islamophobic in his manifesto; he even said that he had no problems with Muslims or Jews as long as they were living in their native lands. He called the disproportion between the birth rates of 'whites' and 'immigrants' as the main cause of worry.

Tarrant's manifesto and live video was that both of these glorified violence as well as the perpetrator like Breivik or Dylan Roof.[262] He called himself an ethno-nationalist and eco-fascist working for ethnic autonomy (all white are a single ethnic community of Europe) of all people and the preserver of 'natural order'.[263] The other two similarities visible in his manifesto with our strategy sections are in terms of religious stereotyping of Muslims as a community with high birth rate and their dehumanization in form of an enemy and an invader.

Tarrant's manifesto is a call to action at all levels. He openly calls for invaders and enemies to be killed justifying it as a need to save themselves from the invasion of immigrants and their increasing population. He said in his manifesto that even a child of the "enemy" will go on to become an adult and produce more invaders. He urges for economic boycott mentioning that white people should not pay taxes to non-white people.

He does not demand for extreme state action as he argues that state is not strong to carry it out and therefore there is a need for someone like him to do something. The social boycott call is visible in his argument for ethno-nationalism focusing on diversity as a weak idea leading to internal conflict.

Tarrant's manifesto is based on an idea of collective identity of a white European person. He answers in his manifesto repeatedly that even though he is an Australian, his European ancestry is enough to support his European similarity[264]. He said that a person like him living in Australia or for that matter someone living in Bavaria makes no difference since both have the similar culture hence focusing on the idea of collective identity and homogenizing.

He calls all European people white and considers Europe should be only for whites therefore calling all other ethnicities invaders and outsiders. The internal moral policing is visible in his manifesto where he talks about who is really to be blamed and he answers, 'ourselves' in terms of European men who have allowed their culture to be degraded.

In his manifesto narrative of blame is visible in Tarrant's idea - he blamed immigrants for high birth rate and an impending 'white genocide'. He blamed immigrants and Muslims for contemporary events by giving news links of rapes of European women at the hands of immigrants. He blamed Muslims for historical events in terms of conversion of Hagia Sofia into a mosque and went on to give a call

[257]Christchurch Call. (2019). Retrieved from https://www.christchurchcall.com/call.html. [18 September 2020].

[258]Ibid.

[259]Owen, T. (16 March 2019). Decoding the racist memes the alleged New Zealand shooter used to communicate. Vice. Retrieved from https://www.vice.com/en/article/vbwn9a/decoding-the-racist-memes-the-new-zealand-shooter-used-to-communicate. [16 October 2020].

[260]Tarrant, B. (2019) The Great Replacement. Retrieved from https://www.docdroid.net/48JypPr/the-great-replacement-original-by-brenton-tarrant-pdf. [14 September 2020].

[261]Tarrant, B. (2019) The Great Replacement. Retrieved from https://www.docdroid.net/48JypPr/the-great-replacement-original-by-brenton-tarrant-pdf. [14 September 2020].

[262]Associated Press. (17 April 2017). Charleston church shooter Dylann Roof pleads guilty to state murder charges. The Guardian. Retrieved from https://www.theguardian.com/us-news/2017/apr/10/charleston-church-shooter-dylann-roof-pleads-guilty-murder-charges. [15 September 2020].

[263]Tarrant, B. (2019) The Great Replacement. Retrieved from https://www.docdroid.net/48JypPr/the-great-replacement-original-by-brenton-tarrant-pdf. [14 September 2020].

[264]Tarrant, B. (2019) The Great Replacement. Retrieved from https://www.docdroid.net/48JypPr/the-great-replacement-original-by-brenton-tarrant-pdf. [14 September 2020].

for action by saying, "We will kill you and drive you roaches from our lands. We are coming for Constantinople and we will destroy every mosque and minaret in the city."[265]

The blame for appeasement is visible in Tarrant's ideas as he blamed NGOs for creating a sense of guilt in European whites and stripping their culture to develop the competitors. He called them traitors responsible for invasion. Tarrant argues in his manifesto that the numbers do not mean anything, white people may be in majority but they will be wiped out because of high birth rate of immigrants.

He himself said, "minorities are never treated well, never become one." [266]

Network leadership and Internal equations are visible in Tarrant's case albeit in a little unorganized form since just one person's account and ideas do not give a clear picture unless confirmed by the other perpetrators. The idea of communal violence as community service drives Tarrant's theme as he repeatedly argues the violence is required to save the future white generations. The terminology calling Breivik, a Knight and similar term used for Tarrant after the arrest in 8chan groups, gives a sense that people like him or Breivik are perceived as leaders. Tarrant argued that while he did not carry out attack with any outside support but he had interacted with many 'national groups' and said that there are millions of soldiers like him.

On the other hand, in terms of Internal Equation, Tarrant claimed that he had been in contact with Breivik. Tarrant's direct aim was to inspire more people and more such attacks. Tarrant inspired more such attacks in next few months as he was glorified as hero on numerous messaging boards like 8chan.[267] A few days later a mosque fire in Escondido, California was connected to Christchurch attack when graffiti related to the attack was found on the site.[268] On 27 April 2019, a shooting occurred at a Synagogue in Poway, California. The attacker cited Tarrant as an inspiration.[269] Tarrant had mentioned Texas specifically as a center of invasion by immigrants and soon enough on 3 August 2019, a 21 year old gunman Patrick Crusius carried out a shooting in El Paso, Texas and killed 23 people.[270]

He followed the similar path of writing a manifesto and posting it on 8chan and cited Tarrant as the inspiration behind the attack.[271] On 10 August 2019 a mosque in Norway was attacked in a similar fashion with the attacker trying to live stream the attack, he cited Tarrant as his inspiration as well.[272] A clear indication of the Christchurch attack acted as an inspiration for many other through a sense of urgency for the similar cause was visible in all attacks afterwards. Tarrant himself had been inspired by Breivik as stated above. The internal equation visible in these cases is similar to the tone of our ethnography.

Tarrant had said that he needs to stop 'shitposting' and do something in real life[273]. In 2020 during Black Lives Matter (BLM) movement, in Kenosha, Wisconsin, a shooting occurred in the name of defense. The accused had posted a live video before the attack and during the protest he was visible with other armed people in the presence of police.[274] The role of Facebook groups and members coming in contact through the suggestion feature was criticized.[275]

---

[265]Ibid

[266]Ibid

[267]Maley, P. (5 September 2019). Accused Christchurch mass killer Brenton Tarrant emerges as far right extremist 'hero'. The Australian. Retrieved from https://www.theaustralian.com.au/world/accused-christchurch-mass-killer-brenton-tarrant-emerges-as-far-right-extremist-hero/news-story/c8c2db7861252527f4587026ff05ab63. [16 September 2020].

[268]Johnson, A. (24 March 2019). Suspect of possible arson attack at Escondido Mosque leaves note referencing New Zealand terrorist attacks". NBC.Retrieved from https://www.nbcsandiego.com/news/local/islamic-center-escondido-mosque-epd-efd-sdso-reported-arson-unit/81831/. [15 September 2020].

[269]Evans, R. (28 April 2019). Ignore the Poway Synagogue shooter's manifesto: Pay attention to 8chan's /pol/ Board. The Belling Cat. Retrieved from https://www.bellingcat.com/news/americas/2019/04/28/ignore-the-poway-synagogue-shooters-manifesto-pay-attention-to-8chans-pol-board/. [15 September 2020].

[270]Romero, S., Fernandez, M., & Padilla, M. (3 August, 2019). Day at a shopping center in Texas turns deadly. The New York Times. Retrieved from https://www.nytimes.com/2019/08/03/us/el-paso-walmart-shooting.html. [15 September 2020].

[271]Embury-Dennis, T. (4 August, 2019). El Paso shooting suspect 'espoused racist tropes and voiced support for Christchurch mosque gunman. The Independent. Retrieved from https://www.independent.co.uk/news/world/americas/el-paso-shooting-suspect-patrick-crusius-white-supremacist-trump-texas-walmart-a9038611.html. [16 September 2020].

[272]Burke, J. (11 August 2019). Norway mosque attack suspect 'inspired by Christchurch and El Paso shootings'. The Guardian. Retrieved from https://www.theguardian.com/world/2019/aug/11/norway-mosque-attack-suspect-may-have-been-inspired-by-christchurch-and-el-paso-shootings. [15 September 2020].

[273]ADL. (2019) Gab and 8chan: Home to terrorist plots hiding in plain sight . Retrieved from file:///C:/Users/DEF/Downloads/gab_and_8chan__home_to_te-14289.pdf. [15 September 2020].

[274]Willis, H., Xiao, M., Tribert, C., Koettl, C., Cooper, S., Botti, D., Ismay, J., & Tiefenthäler, A. (27 August 2020). Tracking the suspect in the fatal Kenosha shootings. The New York Times. Retrieved from https://www.nytimes.com/2020/08/27/us/kyle-rittenhouse-kenosha-shooting-video.html. [15 September 2020].

[275]McEvoy, J. (4 August 2020). Study: Facebook allows and recommends white supremacist, anti-semitic and QA non groups with thousands of members. Forbes. Retrieved from https://www.forbes.com/sites/jemimamcevoy/2020/08/04/study-facebook-allows-and-recommends-white-supremacist-anti-semitic-and-qanon-groups-with-thousands-of-members/#294159e86bbd. [15 September

Tarrant in his manifesto called South Africa leader Nelson Mandela responsible for white genocide and claimed that he will serve 27 years in jail like Mandela "for the same crime" and then receive Nobel peace prize like Mandela received[276]. The theme of centralized conspiracy by Muslims or the Jews and a need of violence to defend the homeland and the culture are visibly similar in Tarrant and inspired attackers' ideas, particularly in resistance to BLM movement. A pattern can be seen although the context visibly differs depending on the geographic locations. Tarrant supported the idea of non-white defending their home land in the way he did. The idea that assimilation of cultures has failed and diversity can be insured only through people living within their respective diverse background is the only way forward.

Tarrant blamed communists as well and called for their death in his manifesto as he argued that they were anti-white[277] The themes and use of social media in terms of strategies, content and call to action is similar in various aspects between Brenton Tarrant and other lone wolf shooters and the patterns observed in our ethnography. These seem to indicate a similarity in terms of use of social media in building the narrative against specific communities.

Common practices

These comparatives allow for an informed perspective on the different ways in which social media can catalyse offline violence. A prominent commonality observed – although it manifested differently depending on socio-political context – is that there is a given subjectivity being tapped into, to strengthen primordial identities and associate claims[278].

An example of how ideas of such identity 'under threat' is produced, looked at through the lens of 'fear of small numbers[279], is common in all three cases discussed here. The common narrative of high birth rate of Muslims and immigrants is a tool that enables majority to think that they are in danger of becoming culturally and numerically in the minority[280] thereby requiring an urgent need to act against this danger.

Apart from recognizing the pervasiveness of exclusionary speech, it becomes equally necessary to identify that there are multiple locations of power behind the picture of calculated populism painted through the content shared within the Facebook groups and pages studied in this research. Facebook's own delayed cognisance of algorithmic design that profits from increased user engagement even if it is hate-fuelled or is inciting offline action,[281] coupled with content moderation that is unevenly applied, exists alongside vested interests, and intentionality that stands to gain from a population rendered devoid of sympathy or empathy towards people of a particular community[282]. Incidentally, in the case of the Christchurch shooting as well as the mass violence in Myanmar and Sri Lanka, the targeted community has been Muslims. The religious hatred and related calls to action studied in this research has also been anti-Muslim.

In his essay on the Radio Télévision Libre des Mille Collines (a Rwandan radio station known for significant contribution in inciting hatred against the Tutsi population) and the creation of a democratic alibi, historian Jean-Pierre Chrétien points out the fallacy in assuming a majority-elected leader is proof of democratic culture while other factors such as human rights, respect for minorities, refusal to acknowledge the exclusion of communities, rule of law, social justice are all considered ancillary.[283] He describes the germination of the extremist propaganda that enabled the Rwandan genocide as set within a traditional socio-racial policy that had been refined for a generation, one that was centred around uniting the Hutu masses around a so-called 'Hutu Power' movement – thus facilitating recruitment and expansion.

2020].

[276]Tarrant, B. (2019) The Great Replacement. Retrieved from https://www.docdroid.net/48JypPr/the-great-replacement-original-by-brenton-tarrant-pdf. [14 September 2020].

[277] ibid.

[278]Bačová, V. (1998). The construction of national identity - on primordialism and instrumentalism. Human Affairs. 8. 29-43. Retrieved from https://www.researchgate.net/publication/266484743_THE_CONSTRUCTION_OF_NATIONAL_IDENTITY_-_ON_PRIMORDIALISM_AND_INSTRUMENTALISM. [17 September 2020].

[279]Sayrah, A. D. (05 May 2018). Facebook helped foment anti-Muslim violence in Sri Lanka. What now? | Amalini De Sayrah. The Guardian. Retrieved, from https://www.theguardian.com/commentisfree/2018/may/05/facebook-anti-muslim-violence-sri-lanka [14 September 2020].

[280]Appadurai, A. (2006). Fear of Small Numbers. In Fear of small numbers an essay on the geography of anger, pp. 49-85. Duke University Press:Durham.

[281]Askonas, J. (Winter 2019). How tech utopia fostered tyranny. The New Atlantis. Retrieved from https://www.thenewatlantis.com/publications/how-tech-utopia-fostered-tyranny#:~:text=The%20emerging%20soft%20authoritarianism%20in,technologies%20toward%20their%20true%20ends.&text=Jon%20Askonas%2C%20%22How%20Tech%20Utopia,57%2C%20Winter%202019%2C%20pp. [17 September 2020].

[282]Laub, Z. (7 June 2019). Hate Speech on Social Media: Global Comparisons. Council On Foreign Relations. Retrieved from https://www.cfr.org/backgrounder/hate-speech-social-media-global-comparisons [18 September 2020].

[283]Chrétien J. (28 December 2010). RTLM Propaganda: the democratic alibi. Retrieved from https://francegenocidetutsi.org/RTLM-propagandaDemocraticAlibi28December2010.pdf [11 September 2020].

This is how, according to Chrétien, democratic language transforms into a technology designed for totalitarian mobilization, under the guise of free speech – the slow and steady creation of a democratic alibi, which dictates that any anger expressed in this form becomes 'democratic anger'[284]. Democratic language, when contextualized with democratized means of cultural/ideological production (i.e. the interrelated networks of closed messaging apps and social media platforms), allows for a more pervasive and customizable spread of propaganda or inflammatory speech, facilitating a more effective way of controlling public mood settings at a hyperlocal, grassroots level.

[284]Ibid.

EVALUATING FACEBOOK'S CONTENT MODERATION POLICY

Content moderation policies as a subject of anthropological articulation – and not just technological implementation – reveals them to be sites of overlap and fluctuation between multiple forces.

Today's heavily mediatized world means that a large part of societal interaction takes place under a hybridized form of governance of legal and technological rationality co-constituting each other, which are themselves formed by specific sociological and cultural contexts.[285]

A social media company that operates on the scale of usage that Facebook does, thereby 'governing' what is essentially the public sphere as it exists today, encodes its own version of arbitration of rights into existence through its self-regulatory mechanisms like content moderation policy.

Facebook's content moderation is a two-step process relying on a combination of algorithmic and human analysis: content deemed as violating community standards gets flagged by users as well as artificial intelligence (AI), after which it is sent to human content moderators to be reviewed and either allowed to remain online or resolved[286].

Facebook's content moderation policy explained in a flowchart [287]

Content moderators tend to be outsourced workers and not full-time employees of the social media company[288]. India, Philippines and Ireland host major hubs of content moderation for Facebook, with other small centres in countries such as Kenya and Latvia which handle content from those regions[289].

External content moderators tend to be poorly paid and their working conditions have been the subject of a lawsuit against Facebook in 2019 which led to a settlement of $52 million in 2020[290].

EXISTING CRITIQUES OF FACEBOOK'S CONTENT MODERATION POLICIES

Many aspects of Facebook's content moderation have been called into question in the last few years, the real-life impacts of which have been felt in terms of increased polarization, offline violence, election interference, and political/ electoral manipulation in several countries[291].

Facebook acknowledged its potential role in the escalation of the violence in Sri Lanka and Myanmar. While in Sri Lanka, Facebook eventually took many of the contentious posts down, they had already been viewed and widely circulated by then[292].

Centre for Policy Alternatives, a Sri-Lankan research and advocacy group, told Facebook representatives about 20 hate groups targeting women and minorities in 2014, in a "detailed research paper that contained dozens of links and screenshots". However, by the end of March 2018, 16 out of the 20 groups were still on Facebook[293].

The Facebook 'groups' feature has received its fair share of criticism in terms of complicity in creating seemingly hidden yet continuously growing spaces that can work as incubators of hatred towards vulnerable groups. A 2016 research study by Facebook researcher and sociologist Monica Lee revealed that 64% of people who had joined an extremist group on the platform did so because the group was promoted by Facebook's automated recommendation tools[294].

---

[285]Gillespie, T. (2018). Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media. Yale University Press: New Haven and London.

[286]Barrett, P. (2020). Who Moderates the Social Media Giants? A Call to End Outsourcing. NYU Stern Center for Business and Human Rights. Retrieved from https://bhr.stern.nyu.edu/tech-content-moderation-june-2020 [25 September 2020].

[287]Ibid.

[288]Ibid.

[289]Ibid

[290]Newton, C. (May 12 2020). Facebook will pay $52 million in settlement with moderators who developed PTSD on the job. The Verge. Retrieved from https://www.theverge.com/2020/5/12/21255870/facebook-content-moderator-settlement-scola-ptsd-mental-health [28 September 2020].

[291]Canales, K. (15 September 2020). A fired Facebook employee wrote a scathing 6,600-word memo detailing the company's failures to stop political manipulation around the world. Business Insider. Retrieved from https://www.businessinsider.in/tech/news/a-fired-facebook-employee-wrote-a-scathing-6600-word-memo-detailing-the-companys-failures-to-stop-political-manipulation-around-the-world/articleshow/78116634.cms [28 September 2020].

[292]Ibid.

[293]Rajagopalan, M. & Nazim, A. (2018). "We had to stop Facebook": When anti-Muslim violence goes viral. Buzzfeed news. Retrieved from https://www.buzzfeednews.com/article/meghara/we-had-to-stop-facebook-when-anti-muslim-violence-goes-viral [19 October 2020].

[294]Horwitz, J. & Seetharaman, D. (26 May 2020). Facebook Executives Shut Down Efforts to Make the Site Less Divisive. Wall Street Journal. Retrieved from https://www.wsj.com/articles/facebook-knows-it-encourages-division-top-executives-nixed-solutions-11590507499 (27 September 2020)

It showed that algorithmically suggested groups and Related Pages suggestions, i.e. the platform's "Groups you should join" and "Discover" algorithms, actually accentuate the problem of extreme speech with filter bubbles. There have also been reports of Facebook's recommendation algorithms actively promoting Holocaust denial pages and groups[295].

In May 2020, the Wall Street Journal reported that Facebook executives ignored the results of its own discovery that its algorithms breed divisiveness, and reasoning given by its policy chief Joel Kaplan was that efforts to make conversations online more 'civil' is a paternalistic approach and should be avoided on those grounds[296].

There was mention of an "an officewide rule to approve any post if no one on hand can read the appropriate language", attributing it as a likely contributor to the violence in Sri Lanka and Myanmar, where harmful posts were routinely allowed to stay up[297]. Facebook's lack of Burmese speaking content moderators was attributed as a key contributor for its failure to respond to violence[298]. In 2015, it only had 2 content moderator who spoke the local language[299].

Facebook deals with the problem of scale through its reliance on automation as a key component of content moderation. However, as the phenomenon of automated decision-making has become more prevalent internationally, so has criticism of its ingrained biases and lack of nuanced understanding of context and social realities.

In terms of algorithmic detection of hate speech, it is widely recognized as inadequate if it cannot comprehend all the languages a country like India has, as well as region and context-specific nuance the way that human content moderators can. Apart from that, Facebook's newsfeed and recommendation algorithms have also been called out for breeding polarization or promoting hate[300].

The Indian context

Existing reports and articles describing different instances and patterns of hate speech and religious polarization through misinformation/disinformation include but are not limited to Avaaz[301], Equality Labs[302], and Caravan[303] prominently among others.

According to the Equality Labs study, 93% of the 1000+ posts it reported to Facebook were not removed at all. This includes content advocating violence, bullying and use of offensive slurs, and other forms of Facebook's Tier 1 hate speech, reflecting a "near total failure of the content moderation process"[304]. The study also states that nearly half of the hate speech Facebook initially removed was later restored, and that 100% of these restored posts were Islamophobic in nature.

According to a report released by Avaaz, as of 19 September 2019, Facebook had acted only upon 96 of the 213 potential breaches of its Community Standards that they flagged.[305] Avaaz also reported that they asked Facebook to deploy a team specifically tasked with proactive monitoring by human content moderators of hate speech in Assam, given that it was (and continues to be) a sensitive period for minorities in the state – and Facebook "refused to commit" to this[306].

[295]Martin, A. (17 August 2020). Facebook algorithm 'actively promoting' Holocaust denial, report warns. Sky News. Retrieved from https://news.sky.com/story/facebook-algorithm-actively-promoting-holocaust-denial-report-warns-12050899?dcmp=snt-sf-twitter (27 September 2020)

[296]Horwitz, J. & Seetharaman, D. (26 May 2020). Facebook Executives Shut Down Efforts to Make the Site Less Divisive. Wall Street Journal. Retrieved from https://www.wsj.com/articles/facebook-knows-it-encourages-division-top-executives-nixed-solutions-11590507499 (27 September 2020)

[297]Fisher, M. (27 December 2018). Inside Facebook's Secret Rulebook for Global Political Speech. New York Times. Retrieved from https://www.nytimes.com/2018/12/27/world/facebook-moderators.html [27 September 2020].

[298]Solon, O. (16 August 2018). Facebook's failure in Myanmar is the work of a blundering toddler. The Guardian. Retrieved from https://www.theguardian.com/technology/2018/aug/16/facebook-myanmar-failure-blundering-toddler [19 October 2020].

[299]Ibid.

[300]Wong, J. C. (2017). How Facebook groups bring people closer together – neo – Nazis included. The Guardian. Retrieved from https://www.theguardian.com/technology/2017/jul/31/extremists-neo-nazis-facebook-groups-social-media-islam [19 October 2020].

[301]Avaaz (October 2019). Megaphone for Hate: Disinformation and Hate Speech on Facebook During Assam's Citizenship Count. Retrieved from https://avaazpress.s3.amazonaws.com/FINAL-Facebook%20in%20Assam_Megaphone%20for%20hate%20-%20Compressed%20(1).pdf [28 September 2020].

[302]Equality Labs. (2019). Facebook India - Towards a Tipping Point of Violence Caste and Religious Hate Speech. Retrieved from https://www.equalitylabs.org/facebook-india-report [28 September 2020].

[303]Pennu, I. (30 January 2019). Are Facebook's community guidelines selectively policing anti-government content?. The Caravan. Retrieved from https://caravanmagazine.in/technology/facebook-community-guidelines-algorithms-target-anti-government-content [22 September 2020].

[304]Equality Labs. (2019). Facebook India - Towards a Tipping Point of Violence Caste and Religious Hate Speech. Retrieved from https://www.equalitylabs.org/facebook-india-report [28 September 2020].

[305]Avaaz. (October 2019). Megaphone for Hate: Disinformation and Hate Speech on Facebook During Assam's Citizenship Count. Retrieved from https://avaazpress.s3.amazonaws.com/FINAL-Facebook%20in%20Assam_Megaphone%20for%20hate%20-%20Compressed%20(1).pdf [28 September 2020].

[306]Ibid.

The Caravan article illustrated political motivations behind company policies of suspending certain accounts as opposed to others, allegedly those that went against the ruling regime[307]. The author of the article further notes how her flagging of the gaps in Facebook's content moderation policies went unheard by the company[308].

A closer look at Facebook's existing vulnerabilities

Apart from inconsistent or inadequate application of the platform's existing content moderation policy i.e. community guidelines, other types of vulnerabilities stand encoded within the affordances of the technology itself. Facebook Lives and private groups have both been called out by Indian journalists/ activists for being instrumental to the virality of content that either called for or glorified violence against a vulnerable community.[309]; [310]

Facebook's Head of Strategic Responses, Neil Potts, claimed to have developed tools to moderate Live Stream videos, when a wave of suicide and self-harm videos were started being posted soon after the feature's launch. This moderating tool allows the moderators to "see user comments on live videos", it also has "an advanced playback speed and replay functionality with an added timestamp to user-reported content and text transcripts to live video with a 'heat map' of user reactions to display the times in a video viewer are engaging with it". [311]

However, in the context of our findings, we did not see Facebook Live videos of violent speech getting taken down often, or the comments on it being removed if they were reflective of hate speech or incitements to violence. The only evidence of moderation witnessed was that certain individual profiles or pages that have large followings and frequently post Live videos promoting violence, sometimes mention while they are live that they are getting notifications of several users start reporting their videos. However, this has only happened with very few of the individual profiles.

Live videos are often used by perpetrators as creating their visual archives of proof and glorification of the act. However, live videos also become the vehicle for propagating dangerous speech (such as advocating mass violence, sexual assault towards religious minorities), carrying enough potential to precipitate further violence against targeted communities, and calls to action.

Apart from private groups that were observed to have higher quantities of posts that qualify as hate speech or incitements to violence as compared to public groups, Facebook also has 'secret groups' which are not visible at all to outsiders; not even their names turn up in searches. These can only be joined through invitation by a current member. Like private groups, posts are less likely to get flagged by users due to being shared in a likeminded community, therefore the tracking and removal of hate speech or calls to violence would largely depend upon Facebook's own mechanisms to identify violations of community guidelines.

In 2019, there were reports of secret groups on Facebook with US military sharing offensive messages about the deaths of migrants in US custody,[312] which sparked a conversation about Facebook's content moderation suffering a blow after the platform announced a 'pivot to privacy' earlier the same year[313].

This 'pivot' was essentially an act of shifting the conversation to end-to-end encryption and the value of private messaging, which was described as a bid to keep users and investors happy while sidestepping the burden of engaging with Facebook's growing content moderation problems[314];[315].

Apart from Facebook Live and private groups, the phenomenon of cross posting or cross pollination of content from different media platforms also

[307]Pennu, I. (30 January 2019). Are Facebook's community guidelines selectively policing anti-government content?. The Caravan. Retrieved from https://caravanmagazine.in/technology/facebook-community-guidelines-algorithms-target-anti-government-content [22 September 2020].

[308]Ibid.

[309]Pennu, I. (30 January 2019). Are Facebook's community guidelines selectively policing anti-government content?. The Caravan. Retrieved from https://caravanmagazine.in/technology/facebook-community-guidelines-algorithms-target-anti-government-content [22 September 2020].

[310]Scroll Staff. (30 December 2019). Delhi: Former constable threatens to shoot anti-CAA protestors, arrested after video goes viral. Scroll. Retrieved from https://scroll.in/latest/948271/delhi-former-constable-threatens-to-shoot-anti-caa-protestors-arrested-after-video-goes-viral [28 September 2020].

[311]Koebler, J., & Cox, J. (23 August 2018). Here's How Facebook Is Trying to Moderate Its Two Billion Users. The Vice. Retrieved from https://www.vice.com/en/article/xwk9zd/how-facebook-content-moderation-works (24 September 2020)

[312]Ortutay, B. (3 July 2019). AP Explains: How Facebook handles speech in 'secret' groups. AP News. Retrieved from https://apnews.com/db6241f527a24dcf924e64089ca30137 [29 September 2020].

[313]Tobin, A. (2 July 2019). Civil Rights Groups Have Been Warning Facebook About Hate Speech In Secret Groups For Years. Pro-Publica. Retrieved from https://www.propublica.org/article/civil-rights-groups-have-been-warning-facebook-about-hate-speech-in-secret-groups-for-years [29 September 2020].

[314]Ibid.

[315]Rothman, M. (6 March 2019). Mark Zuckerberg Announces Facebook's Pivot to Privacy. New Yorker. Retrieved from https://www.newyorker.com/news/current/mark-zuckerberg-announces-facebooks-pivot-to-privacy [29 September 2020].

complicates the process of moderation, i.e. when the Facebook post in question is simply a link to another website which has to be visited for the actual 'content' to be seen/heard.

However, with increasing emphasis on inflammatory content filtering through to violence, has led to the taking down of content praising or justifying hate crime. For example, after a white supremacist killed a protester in Charlottesville, USA in 2017, Facebook reportedly gave training materials to content moderators.

These included marked posts such as "James Fields did nothing wrong" alongside an article on the subject from a conservative website, as an example of content "praising hate crime," and that and others like it "should be removed"[316].

However, in the observations there were significant examples of content that calls for identity-based exclusion, violence, and proof of offline intimidation and threat that were allowed to remain on the platform. This is amplified through repetition by public figures with significant following and an active engaged audience.

A 2019 report by Citizens Against Hate states that despite numerous reports with devastating findings of the kind of speech that is allowed to accumulate and spread online, if there is lack of action by people in power it means that laws exist in a vacuum[317].

In a similar strain, NYU Stern's report quotes Mark Zuckerberg as stating that users bear primary responsibility for policing Facebook – a suggestion that "directly obscures that he and his business colleagues designed the system, flaw and all, and failed to anticipate how much harmful content Facebook would attract"[318].

[316]Cox, J. (25 May 2018). Leaked Documents Show Facebook's Post-Charlottesville Reckoning with American Nazis. Vice. Retrieved from https://www.vice.com/en_us/article/mbkbbq/facebook-charlottesville-leaked-documents-american-nazis [29 September 2020].

[317]Citizens Against Hate. (01 March 2020) Majoritarian Consolidation: Chronicling the Undermining of the Secular Republic. Retrieved from http://citizensagainsthate.org/wp-content/uploads/2020/03/Citizens-Against-Hate-Chronicling-Majoritarian-Consolidation.pdf [29 September 2020].

[318]Barrett, P. (June 2020). Who Moderates the Social Media Giants? A Call to End Outsourcing. NYU Stern Center for Business and Human Rights. Retrieved from https://bhr.stern.nyu.edu/tech-content-moderation-june-2020 [25 September 2020].

EVALUATING THE APPLICATION OF FACEBOOK'S COMMUNITY GUIDELINES[319]

[319]Referenced from https://www.facebook.com/communitystandards/introduction

Relevant sections and subsections of Facebook's community standards page are highlighted and described along with observed patterns of content left unmoderated.

Note: Examples given are reflective of speech patterns observed not directly representative in terms of quantity.

I. Violent and criminal behaviour

Violence and incitement

The relevant sections of the guidelines state the prohibition of:

- Threats that could lead to death, i.e. "Statements of intent to commit high-severity violence; Calls for high-severity violence, including content where no target is specified but a symbol represents the target and/or includes a visual of an armament to represent violence; or Statements advocating for high-severity violence; or Aspirational or conditional statements to commit high-severity violence"

- Threats that could lead to serious injury, i.e. "Admissions, statements of intent or advocacy, calls to action including Statements of intent to commit violence; Statements advocating violence; or Calls for mid-severity violence, including content where no target is specified but a symbol represents the target; or Aspirational or conditional statements to commit violence."

- Threats that could lead to physical harm or other forms of lower-severity violence) towards private individuals (self-reporting required) or minor public figures.

- Misinformation and unverifiable rumours that contribute to the risk of imminent violence or physical harm.

- Statements of intent or advocacy, calls to action, or aspirational or conditional statements to bring weapons to locations, including, but not limited to, places of worship, educational facilities or polling places (or encouraging others to do the same).

Observations: Observations covered 7944 posts, 1898 contained calls to action out of which 637 posts carried direct call to violence, 397 carried call for economic boycott and 225 carried call for social boycott, both of them can be understood as structural violence and another 639 posts carried call for state violence against specific community

Dangerous individuals and organizations

A hate organisation is defined as: Any association of three or more people that is organised under a name, sign or symbol and that has an ideology, statements or physical actions that attack individuals based on characteristics, including race, religious affiliation, nationality, ethnicity, gender, sex, sexual orientation, serious disease or disability.

Observations: Observations included leaders of regional organizations with specific agenda. These organizations at different times have called for action, supported violence against specific communities and the people and have conducted their own raids to discipline and punish. Some have distributed flags to vendors for the purpose of profiling in order to carry out the economic boycott of Others.

A number of such leaders and organizations have targeted individuals, public leaders on their Facebook account through live videos, urging their followers to do the same and filing FIRs and giving threats. They have also urged for extreme state action against specific communities on their social media accounts. Dangerous individual could be seen in two categories, the first categories involved people who were part of organizations explained above and second who were part of opinion channels which normalized the activities of people from first category and justified the violence with the rationale for retributive justice.

Coordinating harm and publicizing crime Harm against people

- Depicting, admitting to or promoting the following acts committed by you or your associates:

- Acts of physical harm against humans, including acts of domestic violence, except when shared in a context of redemption or defence of self or another person

- Statements of intent, calls to action, representing, supporting, or advocating for, or depicting, admitting to or speaking positively about, the following acts committed by you or your associates:

- Swatting

- Depicting, promoting, advocating for or encouraging:

- Participation in a high-risk viral challenge

Harm against property

- Statements of intent, calls to action, representing, supporting or advocating for harm against property that depicts, admits to or promotes the following acts committed by you or your associates:

- Vandalism

Observations: This category of guidelines was visibly broken many times through posts that celebrated violence, escalating particularly in events of public crises. A number of such posts praised the violence against Others for perceived cultural transgressions which escalated around events of public crises. The posts supported and glorified the demolition of important cultural sites. Vandalism was supported in many posts if it was about properties of Others as was cultural regulation regarding use of public spaces.

II. Safety

Child sexual exploitation, abuse and nudity

Content that depicts participation in or advocates for the sexual exploitation of children, including (but not limited to): Engaging in any sexual activity involving minors. Content (including photos, videos, real-world art, digital content and text) that depicts: Any sexual activity involving minors. Content that depicts child nudity where nudity is defined as: Visible genitalia (even when covered or obscured by transparent clothing). According to Facebook's community guidelines it removes nude images of children irrespective of context to prevent the possibility of other people re-using or misappropriating those images.

Observations: A video depicting alleged rape of a minor. The caption above the video described it as a man from the Other community caught attempting to rape a 3-year-old girl. The video does show a man being pulled off a child where both their undergarments are down, and then being chased and beaten by a gathering of men. The caption and comments describe this as the "reality of all" Others, the danger they all pose to society as a collective and that they must be hunted down. Another video depicted a different incident of an alleged sexual assault on a 4-year-old girl by a middle-aged man with a caption along the same lines. Apart from this there was a video involving child nudity with the aim of demonstrating social and cultural practices that inflict pain/ torture upon children, who as a result of their socialisation grow up to become threats to society.

Sexual exploitation of adults

Facebook recognizes and removes content that depicts threatens or promotes sexual violence, sexual assault or sexual exploitation, i.e. content that displays, advocates for or coordinates sexual acts with non-consenting parties or commercial sexual services, such as prostitution and escort services. In instances where content consists of any form of non-consensual sexual touching, crushing, necrophilia or bestiality, or forced stripping, including: Depictions (including real photos/videos), or Advocacy (including aspirational and conditional statements), or Statements of intent, or Calls for action, or Threatening, soliciting or stating an intent to share imagery, or Admitting participation, or Mocking victims of any of the above.

Observations: Multiple posts or comment sections included open threats with sexually violent language about Other women, including encouraging sexual assault or mocking stereotypes promoting non-consensual sex and alleging rape as an accepted social practice. These stereotypes was extensively used to denote the sexual assault and oppression of women in the community. In one instance a Facebook Live video by a female user had gone viral around an event of public crisis calling for continued sexual assault against women from the Other community accompanied with disparaging slurs. There were also posts stereotyping men of the Other community as predatory and maintaining non-consensual sexual relations with young women or girls in the family. A lot of these directly describe sexual violence to women as being endemic to the community. And how their established social and cultural practices institute violence against women.

Bullying and harassment

Facebook recognises that "bullying and harassment happen in many places and come in many different forms, from making threats to releasing personally identifiable information, to sending threatening messages and making unwanted malicious contact".

This includes malicious targeting through:

- Calling for, or making statements of intent to engage in, bullying and/or harassment; Calling for self-injury or suicide of a specific person or group of people; Attacking them through derogatory terms related to sexual activity (e.g. whore, slut); Posting content about a violent tragedy, or victims of violent tragedies, that includes claims that a violent tragedy did not occur;

- Posting content about victims or survivors of violent tragedies by name or by image, with claims that they are: Acting/pretending to be a victim of an event; Otherwise paid or employed to mislead people about their role in the event

- Threatening to release an individual's private phone number, residential address or email address

- Sending messages that contain the following attacks when aimed at an individual or group of individuals in the thread: Targeted swearing, Calls for death, serious disease, disability, epidemic disease or physical harm, Female-gendered cursing terms when used in a derogatory way

- Target public figures by purposefully exposing them to: For adults: Calls for death, serious disease, epidemic disease or disability, Claims about sexually transmitted diseases, Female-gendered cursing terms when used in a derogatory way, Content that praises, celebrates or mocks their death

- Target private individuals or involuntary minor public figures with: Comparisons to animals or insects that are culturally perceived as intellectually or physically inferior, or to an inanimate object ("cow", "monkey", "potato"); Attacks through negative physical descriptions, Content sexualising another adult, Content that praises, celebrates or mocks their death

Observations: Derogatory language, sexual metaphors, identity-based slurs, and disparaging references were extensively used against Others, disaffiliate members of the in-group, as well as sub-groups with alternate political practices. There have been instances where individuals numbers and details have been made public as a method of intimidation both against public figures as well as private individuals. These speech practices tend to achieve virality around key figures during key public events or crises.

Privacy violations and image privacy rights

Facebook does not allows one to post content that may reveal one's personal information i.e. Content that identifies individuals by name and depicts their personal information, including: Driving licences; government IDs other than driving licences; Green Cards or immigration papers; Marriage, birth and name change certificates; Digital identities, including passwords and number plates.

Observation: There have been instances where the personal information of inter-group couples like their marriage certificates (which carry their names, photographs, addresses) have been shared on the public and private groups with the intent to malign them under the alleged conspiracy of social and cultural erosion. There was also at least one post observed in the private groups that contained details of an inter-group couple staying together at a particular hotel, with the post asking for people to hunt them down as their status as an mixed couple is unacceptable and the girl is likely being wronged in some way.

III. OBJECTIONABLE CONTENT

Hate speech (Tier 1, Tier 2, and Tier 3)

Hate speech is defined as "a direct attack on people based on what we call protected characteristics – race, ethnicity, national origin, religious affiliation, sexual orientation, caste, sex, gender, gender identity and serious disease or disability". The word attack indicates violent or dehumanising speech, harmful stereotypes, statements of inferiority or calls for exclusion or segregation. Attacks are divided into three tiers of severity, as described below:

Tier 1: Content targeting a person or group of people (including all subsets except those described as having carried out violent crimes or sexual offences) on the basis of their aforementioned

protected characteristic(s) or immigration status with:

- *Violent speech or support in written or visual form*
- *Dehumanising speech or imagery in the form of comparisons, generalisations or unqualified behavioural statements (in written or visual form) to or about:*
- *Insects*
- *Animals that are culturally perceived as intellectually or physically inferior*
- *Filth, bacteria, disease and faeces*
- *Sexual predator*
- *Subhumanity*
- *Violent and sexual criminals*
- *Other criminals (including but not limited to "thieves", "bank robbers" or saying "All [protected characteristic or quasi-protected characteristic] are 'criminals'")*
- *Statements denying existence*
- *Mocking the concept, events or victims of hate crimes, even if no real person is depicted in an image*
- *Designated dehumanising comparisons, generalisations or behavioural statements (in written or visual form) that include:*
- *Black people and apes or ape-like creatures*
- *Black people and farm equipment*
- *Caricatures of Black people in the form of blackface*
- *Jewish people and rats*
- *Jewish people running the world or controlling major institutions such as media networks, the economy or the government*
- *Muslim people and pigs*
- *Muslim person and sexual relations with goats or pigs*
- *Mexican people and worm-like creatures*
- *Women as household objects or referring to women as property or "objects"*
- *Transgender or non-binary people referred to as "it"*

Tier 2: Content targeting a person or group of people on the basis of their protected characteristic(s) with:

- *Generalisations that state inferiority (in written or visual form) in the following ways:*
- *Physical deficiencies are defined as those about:*
- *Hygiene, including, but not limited to: filthy, dirty, smelly*
- *Physical appearance, including, but not limited to: ugly, hideous*
- *Mental deficiencies are defined as those about:*
- *Intellectual capacity, including, but not limited to: dumb, stupid, idiots*
- *Education, including, but not limited to: illiterate, uneducated*
- *Mental health, including, but not limited to: mentally ill, retarded, crazy, insane*
- *Moral deficiencies are defined as those about:*
- *Character traits culturally perceived as negative, including but not limited to: coward, liar, arrogant, ignorant*
- *Derogatory terms related to sexual activity,*

including, but not limited to: whore, slut, perverts

- *Other statements of inferiority, which we define as:*
- *Expressions about being less than adequate, including, but not limited to: worthless, useless*
- *Expressions about being better/worse than another protected characteristic, including, but not limited to: "I believe that males are superior to females."*
- *Expressions about deviating from the norm, including, but not limited to: freaks, abnormal*
- *Expressions of contempt (in written or visual form), which we define as:*
- *Self-admission to intolerance on the basis of protected characteristics, including, but not limited to: homophobic, islamophobic, racist*
- *Expressions that a protected characteristic shouldn't exist*
- *Expressions of hate, including, but not limited to: despise, hate*
- *Expressions of dismissal, including, but not limited to: don't respect, don't like, don't care for*
- *Expressions of disgust (in written or visual form), which we define as:*
- *Expressions suggesting that the target causes sickness, including, but not limited to: vomit, throw up*
- *Expressions of repulsion or distaste, including, but not limited to: vile, disgusting, yuck*
- *Cursing, defined as:*
- *Referring to the target as genitalia or anus, including, but not limited to: cunt, dick, asshole*
- *Profane terms or phrases with the intent to insult, including, but not limited to: fuck, bitch, motherfucker*
- *Terms or phrases calling for engagement in sexual activity, or contact with genitalia, anus, faeces or urine, including but not limited to: suck my dick, kiss my ass, eat shit*

Tier 3: Content targeting a person or group of people on the basis of their protected characteristic(s) with any of the following:

- *Calls for segregation*
- *Explicit exclusion, which includes, but is not limited to, "expel" or "not allowed".*
- *Political exclusion defined as denial of right to political participation.*
- *Economic exclusion defined as denial of access to economic entitlements and limiting participation in the labour market,*
- *Social exclusion defined as including, but not limited to, denial of opportunity to gain access to spaces (incl. online) and social services.*

We do allow criticism of immigration policies and arguments for restricting those policies. Content that describes or negatively targets people with slurs, where slurs are defined as words commonly used as insulting labels for the above-listed characteristics.

Observations: All three tiers of hate speech have been observed to be highly prevalent in terms of targeting the Others on the basis of their protected characteristics. The strategies mobilised within speech practices – these include narrativized disinformation campaigns, dehumanizing language including comparisons to cockroaches, snakes, pigs; widespread usage of designated slurs, cursing, and profanity; stereotyping including referring to essentialised characteristics of the men from the community as sexual predators; with the community being essentially that of violent and barbaric people; designating the women as oppressed and inherently subservient to the men but also calling for sexual assault against women from the community, as well sexualising key women figures suggesting them to be of questionable moral character; scapegoating the community during and around events of public crises.

Apart from direct calls for physical violence there were also calls for economic and social exclusion,

for e.g. calls to boycott vendors and traders from the Other community and buy from vendors and traders from one's own group. Towards this end, signifiers and symbols for distributed to enable easy identification such vendors as mentioned above. Even though these are covered under Tier 3 hate speech of Facebook's community guidelines, they work towards creating an enabling environment where violence is normalised or justified. While offline violence takes on different forms, (i) these practices have the capacity to cause direct and tangible harm, and (ii) are interlinked through the speech that is normalizing them, whether it is through demonizing the 'other', or glorifying in-group superiority.

Violent and graphic content

Content that glorifies violence or celebrates the suffering or humiliation of others because it may create an environment that discourages participation. This includes videos of people or dead bodies in non-medical settings if they depict:

- *Dismemberment;*
- *Visible internal organs, partially decomposed bodies;*
- *Charred or burning people unless in the context of cremation or self-immolation when that action is a form of political speech or newsworthy.*

Observations: Glorification of violence against Others has been an important narrative strategy for building collective identity. These also extend to posts captioned as depicting members of the in-group as victims of violence perpetrated by the Other. Specific instances include pictures and videos of dead bodies or grievously injured individuals, shared as victims of violence against the in-group as part either of civic violence or individual cases.

Specific instances of civic violence are re-scripted as having been targeted specifically against the in-group, despite evidence that of a large section of the victims being the Others. This re-scripting is done through narrative manipulation by sharing captioned videos or amplifying affiliative identity of the victims of civic violence. This is translated as the essentialised barbarism perpetrated by the Other and mobilise meaning-making resources towards reinforcing the narrative of the in-group under threat.

Another pattern of violent imagery included posting images or videos of victims of violent sexual assault. One particular post included the video of a man explaining an image of an injured unconscious girl with a lot of blood on the floor. The man tells his audience that the girl in the image is the victim of a gang-rape and her family being under substantial pressure with people attempting to bribe them. The caption to the post signified and emphasized the identities of the both the victims and the 4 alleged perpetrators. Multiple references were made to draw the attention of the viewer to the condition of the girl and the grievous assault suffered by her.

While posts re-scripted to reinforce the affiliative identity of alleged victims of violence, videos of victims from the Other community at the receiving end of state action were met with derisive celebratory response including designated slurs and stereotypes.

Adult nudity and sexual activity

- *Explicit sexual intercourse, defined as mouth or genitals entering or in contact with another person's genitals or anus, where at least one person's genitals are nude*
- *Implied sexual intercourse, defined as mouth or genitals entering or in contact with another person's genitals or anus, even when the contact is not directly visible, except in cases of sexual health context, ads and recognised fictional images or with indicators of fiction.*

Observations: In one of the videos that were observed that aimed at delegitimizing women led civic participation, there was depiction of a sexual intercourse between a man and a woman where cultural signifiers worn by the woman were indications of her identity. The location of the video was uncertain, however, in the background could be heard speeches and statements that signified the location to be within the vicinity of the site of civic participation. The purpose behind the video was to delegimitise the women partaking in the civic participation. The range of captions accompanying the post referred to the women as prostitutes.

Cruel and insensitive

Content that depicts real people and laughs at or makes fun of their serious physical injury, starvation, or serious or fatal disease or disability.

Content that contains sadistic remarks and any visual or written depiction of real people experiencing premature death, serious physical injury, physical violence or domestic violence.

Observations: There were several videos that mocked extreme police action meted out for violating movement restrictions during COVID – 19,

particularly when they were against the members of the Other community. These included instances of police chasing them out of designated cultural locations, dragging them out of homes for testing, beating them with batons, making them perform humiliating tasks like imitating frogs or ducks, or even just beating stragglers on the road. The comment section express derision and support for the action with the use of derogatory, disparaging, and designated slurs and cursing against the victim. An exemplar of how these videos were captioned include, 'live action entertainment'.

IV. Integrity and authenticity

Misrepresentation

Facebook disallows the misuse of their products under the following circumstances:

- Maintaining multiple accounts.
- Creating another Facebook or Instagram account after being banned from the site.
- Creating or managing a Page, group, event or Instagram profile because the previous Page, group, event or Instagram profile was removed from the site.

Observations: It was observed that individuals operate 2 accounts as access to their previous account was blocked by Facebook for a given period of time. The second account was operated as a back-up account when access to their primary account was blocked. Further, new accounts were created when old accounts were permanently taken down by Facebook. These dynamics were well understood by the networked leadership who engaged their audiences in cementing their online presence. In one instance, a page posted a link for a 'back-up' page as the existing one was getting reported for its content. Another tactic that was used involved asking followers to give the pages 5-stars and leave a positive review so that it could be prevented from being taken down.

False news

Reducing the spread of false news on Facebook is a responsibility that we take seriously. We also recognise that this is a challenging and sensitive issue. We want to help people stay informed without stifling productive public discourse. There is also a fine line between false news and satire or opinion. For these reasons, we don't remove false news from Facebook, but instead significantly reduce its distribution by showing it lower in the News Feed." This is achieved mainly by:

- *Using various signals, including feedback from our community, to inform a machine learning model that predicts which stories may be false.*
- *Reducing the distribution of content rated as false by independent third-party fact-checkers.*
- *Empowering people to decide for themselves what to read, trust and share by informing them with more context and promoting news literacy*

Observations: Existing fact-checking procedures fail to take into account hybrid posts involving narrativization, re-contextualising, and re-scripting. Moreover, fact-checking can only debunk false information and not de-bias people[320]. However, even the fact-checking disclaimers have appeared alongside posts, it has usually appeared in English as opposed to the local language of conversation of the group/ page in which such content was circulated. In one instance it appeared in a completely different local language which would have been unreadable for the average audience of such group/ page.

"It was observed that individuals operate 2 accounts as access to their previous account was blocked by Facebook for a given period of time. The second account was operated as a back-up account when access to their primary account was blocked. Further, new accounts were created when old accounts were permanently taken down by Facebook. These dynamics were well understood by the networked leadership who engaged their audiences in cementing their online presence. In one instance, a page posted a link for a 'back-up' page as the existing one was getting reported for its content."

[320]Digital Empowerment Foundation. (2019). Digital Citizen Summit. Retrieved from https://www.defindia.org/wp-content/uploads/2020/06/DCS-Report-2019.pdf [22 October 2020].

RECOMMENDATIONS

The internet and social media technologies continue to have immense emancipatory and democratic potential, and keeping that in mind we do not believe in advocating for blanket increase in censorship and surveillance but rather focusing on processes that allow for greater transparency and accountability.

- **Leveraging civil society experience:** In order to maintain the democratic fabric of the platform as an inclusive right-respecting and empowering space, serious attention needs to be devoted to take civil society research and reports into consideration. This needs to be done through consistent engagement with civil society organizations and activists who are working in a wide range of regions and contexts in order to enable better informed moderation practices.

- **Develop and deploy responsive content moderation trainings, particularly in sensitive contexts:** This relates not just to an increase in the number of moderators with local sensitivity, but also the focus towards rapid development of a more holistic approach that takes into account all regional languages of the country as well contextual templates and situational models for content moderators to follow.

- **Increased decisional transparency:** There should be greater decisional transparency in terms of how moderation works in dealing with harmful content. How moderation decisions and responses are applied to different level of speech and how such categories are applied. In other words, how rules are applied in practice and its process for dealing with reports submitted by external researchers as proofs of unmoderated hate speech as opposed to the process mechanism of user reports on the platform. So far Facebook has not yet achieved what UN Special Rapporteur David Kaye calls 'decisional transparency'[321].

- **Increased procedural transparency:** While community guidelines outline norms of acceptable/ unacceptable behaviour, the mechanism for enforcing them on the platform is not clear. For example, it is not apparent what procedures might be put into motion for violation of community guidelines and how they are tiered according to different categories of violations.

- **Notification for content labelled to be fake:** An ex-post notification for information labelled as fake by fact-checking agencies. Particularly, for pages with large following which have the highest potential for instrumentalised virality for narrativized misinformation.

- **Reviewing recommendation guidelines:** Given Facebook's existing recommendation algorithm is pushing people towards extremist filter-bubbles, it is extremely crucial for Facebook to review the parameters of its existing recommendation algorithms in order to foster inclusive democratic spaces.

[321] Ash, T. G., Gorwa, R., & Metaxa, D. (January 2019). GLASNOST! Nine ways Facebook can make itself a better forum for free speech and democracy. Retrieved from https://reutersinstitute.politics.ox.ac.uk/our-research/glasnost-nine-ways-facebook-can-make-itself-better-forum-free-speech-and-democracy [24 September 2020]

Email: def@defindia.net
URL: www.defindia.org

DIGITAL EMPOWERMENT *foundation*

FUNDING DISCLAIMER

In 2019, the Digital Empowerment Foundation was one of the recipients of Facebook Content Policy Research Awards to understand the linkages between hate speech and offline violence in India.

facebook