International Journal of Communication 11(2017), 1173–1191 1932–8036/20170005

Extreme Speech Online: An Anthropological Critique of Hate Speech Debates

MATTI POHJONEN
Africa's Voices Foundation, UK

SAHANA UDUPA
Central European University, Hungary

Exploring the cases of India and Ethiopia, this article develops the concept of "extreme speech" to critically analyze the cultures of vitriolic exchange on Internet-enabled media. While online abuse is largely understood as "hate speech," we make two interventions to problematize the presuppositions of this widely invoked concept. First, extreme speech emphasizes the need to contextualize online debate with an attention to user practices and particular histories of speech cultures. Second, related to context, is the ambiguity of online vitriol, which defies a simple antonymous conception of hate speech versus acceptable speech. The article advances this analysis using the approach of "comparative practice," which, we suggest, complicates the discourse of Internet "risk" increasingly invoked to legitimate online speech restrictions.

Keywords: online abuse, hate speech, India, Ethiopia, comparative practice, Internet risk

The recent electoral victories for conservative groups with aggressive online presence have brought the political stakes of digital speech into sharp public focus, unsettling euphoric pronouncements on new media as a radical enabler of citizen participation and open society. Whether online Islamist radicalization or hate messages on social media during the 2016 refugee crisis, current developments have reinvigorated political debates on the limits of free speech online. The discourse on digital technologies has tilted toward the "dark side" of new media as a platform for promoting hate speech, fake news, right-wing nationalist mobilization, terrorism, misogyny, and intergroup conflict (Lovink, 2013; Morozov, 2012). Such negative forms of online speech, it is argued, threaten many of the taken-for-granted freedoms commonly associated with digital media cultures around the world. The discourse of online speech as a form of "risk" and "threat" is also used increasingly by governments to rhetorically legitimize securitization and control over their citizens' communicative practices (Amoore & Goede, 2008).

Matti Pohjonen: matti.pohjonen@gmail.com
Sahana Udupa: udupas@spp.ceu.edu
Date submitted: 2016–05–12

[1] Matti Pohjonen's work was partially supported by the European Union's Framework Programme 7 (Grant number 312827: VOX-Pol Network of Excellence).

A cursory glance at policy debates worldwide testifies to the growing anxiety over online content taking negative pathways. In postwar Sri Lanka, studies find that hate messages against Muslims expressed in the local Sinhalese language are more "vicious and venomous" than the anti-LTTE (Liberation Tigers of Tamil Eelam) sentiments even at the peak of the civil war (Samaratunge & Hattotuwa, 2014). The Umati Project in Kenya cautioned against online speech inciting violence, especially during the violent aftermath of the elections in 2007 (Benesch, 2012; Sambuli, Morara, & Mahihu, 2014). Legal provisions to prevent hate speech on grounds of religious harmony and national security are routinely invoked to regulate online media in India, Pakistan, Malaysia, and Sri Lanka. In Europe, widespread racism and xenophobia on Facebook in the wake of refugee arrivals has once again foregrounded the heated debate about the limits of freedom of speech and measures to tackle online aggression (Rowbottom, 2012).

This article shares the concerns raised by the hate speech discourse, but it brings a broader ambit of digital media practices into focus by asking how and under what circumstances do different online actors engage in online vitriol, and with what implications. In so doing, we develop the concept of "extreme speech" as an anthropological qualification for the widely used regulatory term "hate speech." With extreme speech we emphasize practice—that is, what people do that is related to media (Couldry, 2010)—to avoid predetermining the effects of online volatile speech as vilifying, polarizing, or lethal. This entails a departure from assumptions around politeness, civility, or abuse as universal features of communication with little cultural variation—a perspective common in political communication and regulatory debates. We instead gesture toward the situatedness of online speech forms in different cultural and political milieus. This implies two analytical moves.

First, extreme speech recognizes the inherent ambiguity of speech contexts. This ambiguity comes from a tension between the rationale of public interest that defines the contours of speech governance and the interests of publicity that constitute practical politics (Mazzarella & Kaur, 2009). The tension between incitement and containment defines the complex politics behind appropriating and translating the category of hate speech in various national and regional contexts.

Second, extreme speech signals a *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. To approach this as a spectrum forces us to pay attention to online practices that defy easy binary division into speech that is acceptable and speech that is not.

Both the analytical moves require us to place the contemporary moment of online volatile speech within regional and historical context. Such contextualized understanding, we suggest, calls for a comparative analysis that widens the lens beyond the West and turns the focus on the rapidly expanding online worlds of the global South. We take a step in this direction by presenting two important national scenarios of online expansion from the South.[2]

[2] While we are aware of the limits of methodological nationalism (Beck, 2000), we aim to sketch the contours of digital media practice in relation to the regulatory discourse largely articulated at the national level. Although our analysis is rooted in ethnography, for space considerations, we limit our discussion to insights drawn from ethnography without the subtle details of "thick description" (Geertz, 1973). This comparative

In the following discussion, we begin with a critical overview of the key terms and concepts used in understanding online vitriol and position our conceptual intervention of extreme speech within this discursive bundle. We foreground two case studies—India and Ethiopia—first to discuss how hate speech is invoked as a regulatory instrument in these two countries (to illustrate the first analytical move) and then to pry open some of the historical and contextual details of online aggression (the second analytical move). In conclusion, we call for a new critical typology that would unbundle "thick concepts" (Brubaker & Cooper, 2000) such as hate speech and offer a more textured understanding of online vitriol. This, we suggest, can provide conceptual distance from the ongoing discourse of securitization of online speech and the resulting corrosion of online freedoms. The debates around online risk, we argue, are largely devoid of comparative examples from outside the European Union and the United States. A theoretical exposition through the global South could thus help avoid some of the moral panics raised by the emerging focus on the dark side of Internet freedoms.

Online Aggression: Hate Speech to Trolling

In the past decade, the legal-regulatory terminology of hate speech has become an important category in efforts to recognize aggressive speech expanding on online media. This has drawn on the longer legal debates on speech restrictions (Nockleby, 2000; Warner & Hirschberg, 2012). Although differences exist among legal traditions as well as within scholarly discussions, a common element throughout the discourse is that hate speech involves disparagement of other groups based on their belonging to a particular group of collective identity. Waldron (2012) argues that this kind of speech has two key characteristics: The first is to dehumanize members who belong to another group, and the second is to reinforce the boundaries of the in-group against the out-group by attacking the members from the other group. Hate speech discourse predefines the effects of hate speech as negative and damaging, and its regulatory rationale is thus of control and containment. The state is the largest actor in this effort, but Internet intermediaries also increasingly monitor and restrict speech from their platforms. Responding to civil society concerns, governmental injunctions, and international conventions on hate speech, online forums and social networking sites have developed their own terms of service to detect, regulate, and prohibit hate speech. One example is Facebook, which states that "Content that attacks people based on their actual or perceived race, ethnicity, national origin, religion, sex, gender, sexual orientation, disability or disease is not allowed" (Facebook, 2017). But it quickly adds that, "We do, however, allow clear attempts at humor or satire that might otherwise be considered a possible threat or attack. This includes content that many people may find to be in bad taste (ex: jokes, stand-up comedy, popular song lyrics, etc.)." Google YouTube's terms of service admit similarly that there is a "fine line" between what is hate speech and what is not. So, it declares, "It is generally okay to criticize a nation, but not okay to post

---

exercise neither presumes that countries grouped under the global South are similar, nor does it approach digital trajectories beyond the West as cases for normative comparison. On the contrary, it acknowledges the differences in the global discourse such as India's recent framing as an emerging economy versus the more developmental discourse of Ethiopia's digital world through its imagined risks of geopolitical instability and conflict. The purpose here is to locate the variations in online practice and the circulation of global regulatory values as a way to critique the universalizing tendency of the hate speech discourse.

malicious hateful comments about a group solely based on their race" (YouTube, 2017). Twitter's terms of service narrows down the definition to abuse that threatens safety (Wang, 2013).

As it jostles between the state, capitalist market, and realpolitik, hate speech has thus become a "thick concept" with a tangle of different meanings and evaluative load. Hate speech and its online techniques of trolling (Hardaker, 2010) and bullying (Marwick & boyd, 2011) recognize the gravity of hatred and vitriol circulating in cyberspace. As critics point out, they constrict participation in free deliberative dialogue through overt threat, abuse, and stinging stereotypes based on race, gender, ethnic origin, religion, sexual orientation, or nationality. Yet such universalizing concepts risk glossing over a diversity of practices both online and off-line and apply notions derived from the experiences of online media expansion in the West as generalizable models for the rest of the world. Moreover, these concepts become empirical objects in themselves; the task of the researcher would be to merely discover the degree of variance or agreement different kinds of online speech have with this ideal object type. Extreme speech calls into question such contextual flattening.

An influential effort to move the debate toward a more specific understanding of online hatred is recent work that categorizes online speech based on its potential to trigger off-line violence. Benesch's (2012) concept of "dangerous speech" demarcates speech acts that could be precursors to physical violence. Dangerous speech could thus be considered a "thin concept" or "less congested term" (Brubaker & Cooper, 2000). The formulation of "fear speech" proposed by Buyse (2014) maps the sociopsychological dynamics underlying hate speech that would allow detecting situations where the need to mitigate violent content is urgent. Both Benesch and Buyse recognize the importance of understanding the communicative dynamics to distinguish "dangerous" speech from other forms of hate speech. However, these approaches are rooted in globally circulating rights discourse with little room for the cultural dynamics shaping online practices. Extreme speech underscores the need for a thorough ethnographic exploration to grasp how different situational features, including technology, online agency, and political cultures, can lead to various kinds of speech—harmless in some contexts, but with serious political ramifications in others.

The kind of contextual understanding we propose benefits from a comparative analysis. The comparative approach here is not based on a model with quantitative metrics tested across selected case studies, but rather is rooted in ethnography of practice and historical anthropology (van der Veer, 2016). Comparative study of online practice, we suggest, provides a way to unpack concepts such as hate speech and trace the ramifications of appropriating hate speech as a regulatory value in different national-regional contexts. A juxtaposition of India and Ethiopia demonstrates that the diverse agendas behind online media growth represent complex political constellations—an analysis of which might provide a way to complicate the discourse of hate speech.[3]

[3] Although we take up the cases of Ethiopia and India to demonstrate the merits of comparative analysis rooted in ethnography of practice, studies should encompass a breadth of regions, including North–South comparisons, to focus simultaneously on the global features and local specificities of new media.

Online Vitriol in India

As one of the world's fastest growing digital media markets, India's population of 300 million Internet users is next only to China and the United States. A large number of Internet users in India come from the middle class and the well-to-do, but the spread of affordable smartphones in recent years has broadened the class base and narrowed the rural–urban divide in online access. Excitement around new media is evident in the huge uptake for social media networking sites such as Facebook, YouTube, and Twitter and the micro-messaging services of WhatsApp. The ambitious state initiative for Digital India has added further momentum, as private-sector players scurry to make the most of digital expansion. In the midst of new media growth for multiple agendas of development, governance, leisure, and politics, an intriguing practice has caught public attention—the growing invective language and abusive exchange on social media platforms. Online vitriol has recharged public concerns over hate speech, as century-old legal provisions are hastily reworked to address the digital age.

In the Indian legal corpus, explicit mention of hate speech is rare, but restrictions on speech and expression date back to colonial times, when a substantive legal corpus was built around what is now understood as hate speech. Speech regulations are rooted in the colonial state's rationale of law and order, and what it left behind in post-Independence India as the constitutional value of "ordered society" (Narrain, 2016). The colonial regime's apprehensions about indigenous uprising and perceptions of "native" society as prone to "religious excitement" underpin the law-and-order justification for speech regulation (Rajagopal, 2001). As A. Ahmed (2009) states, legal restrictions on speech enunciated a strategy that "enabled the colonial state to assume the role of the rational and neutral arbiter of supposedly endemic and inevitable religious conflicts between what it presumed were its religiously and emotionally excitable subjects" (p. 173).

In the post-Independence years, four key concerns have driven the regulatory and legal action around speech restrictions. First, and by far the strongest, apprehension is around religious difference, which is covered under various penal provisions to address incitement of hatred between religious communities and insult to religion. Tensions between the majority Hindus and minoritized Muslims and Christians mediated by the state, law, and public cultures underwrite a large number of restrictions on speech. This is historically conditioned by the postcolonial politics of religious difference which is inextricably entwined with the professed secular-liberal model of governance (van der Veer, 2001). Restrictions on speech in this line are also reflective of the more recent global discourse around "risk" posed by transborder religious solidarities for national stability. In the online domain, for instance, Indian law enforcement agencies blocked more than 650 online posts and pages in the wake of the *Charlie Hebdo* attacks. It issued takedown notices to Internet intermediaries when images of purported violence against Bangladeshi Muslim migrants in North East India created a wave of panic in major Indian cities in 2012.

Closely related to religious tensions are the provisions rooted in colonial laws around "sedition" to restrict speech that incites disorder and violence endangering national security and stability. The other two realms for speech restriction relate to caste hatred and public representations of sex and sexuality, together with contempt of court and defamation laws. Across these interrelated and crisscrossing domains, not only the postcolonial state but also nonstate actors such as street ruffians and groups acting

on behalf of political parties enact censorship. In many cases, nonstate public cultural interventions have encouraged a heckler's veto legitimizing "vigilante censorship" as an extraconstitutional restriction on speech (Narrain, 2016). Media, and digital media in recent times, are often the locus of explicit and implicit prohibitions.

However, the scenario is far from unilateral repression. The regulatory environment is replete with contradictions. One the one hand are the efforts to involve digital media, as with other forms of media, for state-led developmental goals and economic liberalization for business development (Udupa, 2015). On the other hand, the law-and-order rationale and the painful process of litigation sit in tension with a more liberal jurisprudence around speech restriction. In addition, there are multifarious efforts among political leaders to instrumentalize digital media for propaganda. As a result, hate speech and its cognates have had a checkered course of protection and restriction both within the domain of state-sanctioned institutions and the broader social field.

The tension is evident in recent attempts to curb online speech and intense contestations that followed. In 2015, the Supreme Court of India struck down Section 66A of the Information Technology Act 2000, which criminalized offensive online content, hearing a writ petition that said the provisions violated Article 19 of the Constitution, which guarantees freedom of expression. The Supreme Court of India is hearing a petition (in 2016) by a prominent politician who has challenged several legal provisions to restrict speech, arguing that they overstep the rationale of reasonable restrictions on free expression outlined by the Constitution ("Hate Speech," 2017). Tense exchanges have continued to animate the debate. In a recent report submitted to the Parliament, lawmakers have urged for greater legal provisions to specifically address online content (Parliament of India, 2015). In an elaborate section, the report recommends legal sections to punish online users who send and transmit content "which promotes ill will, hatred and enmity among communities, race, religions etc." (see paras 3.5.4–3.5.9). The report suggests that not only producers of information but also users who forward the content, even when claimed as innocent sharing, should be liable for the offense. The report reasons that only such a provision can act as "a deterrent in the viral spread of such (hate) content," illustrating again the invocation of hate speech for law-and-order problems.

Speech restrictions have been a common feature of governments with different political ideologies, including the current regime representing right-wing politics of religious nationalism and the professedly secular-liberal Indian National Congress party and various regional parties. Unlike Ethiopia, discussed in the next section, hate speech restrictions have been preoccupied more with domestic politics of religious difference and concerns over transborder terrorism than with the influence of international agencies and conventions.

The regulatory scenario reveals the strategies and contestations around a reified category called *hate speech*, but leaves many questions on why and how online vitriol has spread on social networking sites in India. What user practices cohere around online volatility? Can this be understood only as hate speech, and what dangers are there in delineating online abuse as hate speech?

In the cities of Mumbai, Bangalore, and Delhi, where Udupa has been carrying out fieldwork, online users showed repetitive and formatted abuse that clogs their Twitter feeds and Facebook time lines. Comments sections of organized media are similarly filled with online vitriol of swearwords, name-calling and put-downs. Rather than a mere sequence of intentional tit-for-tat actions, abuse frames the context where meanings of political participation are reconfigured for a growing number of online users entering the debate culture of new media. Repetitive trolls revealing the characteristics of bots regularly combine with tweets by real individuals. The category of hate speech obscures more than illuminates this intriguing online culture.

In practice, online abuse cultures have emerged at the junction of technology, market, and political cultures of speech in India. The experiential salience of instantaneity, rapid reaction loops, affordances for relative anonymity, and the possibility to automate trolls and invite attention of interested bystanders through tags and retweets have amplified the conditions for confrontational encounters on online media. Added to this is the premium placed by the market on brief messages to augment data aggregation for consumer analytics and display on small screens (Fuchs, 2015), as with Twitter allowing 140 characters. A political tweeter described how social media has "screwed" his language in such a market-inflected technological milieu.

Historically shaped political cultures of speech and expansion of creative wordplay in political discourse in India in recent years frame the more immediate context. If many of the earlier strategies of language privileged high literary prose, language play started to tilt toward lighter, everyday speech forms at the turn of the millennium. This was a reflection of the popularity of colloquial language use on FM radio and private television, which expanded at a blistering pace after media deregulation in the 1990s. At the same time, political cultures of India are also characterized by "profitable provocation," (p. 3) when political actors willfully breach the "boundaries of public civility and decorum" to gain traction and navigate the "volatile calculus of provocation and respectability, defiance and dignity" (Mazzarella & Kaur, 2009, p. 9). Aside from practical politics, of particular relevance to online abuse is the acceptability of abuse in particular ritual contexts of temple processions and marriage ceremonies as well as routine detoxification of swearwords as a sign of masculine camaraderie in everyday youth cultures (Udupa, forthcoming).

Social media have thus cemented a culture of colloquialism in the political discourse, providing the communicative context for online vitriol to expand. The distinctness of online abusive exchange in India is captured by the emic term *gaali* (in Hindi), which signals the interlocking practices of insult, comedy, shame, and abuse that unfold in a blurred arena of online verbal art. On this slippery ground of shifting terms, comedy stops and insult begins or insult morphs into abuse in mutually generative ways. A striking example for the interplay of the Internet's network architecture, globally shared cyberculture of irreverence, and political practice of abuse in India is the popularity of new online production houses with what is increasingly offered as insult comedy. All India Bhakchod and The Viral Fever are two production groups that regularly parody and satirize social and political realities of urban India with freewheeling flourish of *gaali*, while also, alongside comedy groups such as East India Comedy, strongly advocating the need for free speech (see Rao, 2016).

Beyond the rapidly commercializing low-cost online production groups relying on comedy and parody, online *gaali* has opened new avenues of participation for politically savvy net users. This is especially the case with educated middle-class groups in urban India who feel confident that they can trump legacy media and political authorities with avenues opened up by social media. A good example is the regular exchange of online *gaali* on Twitter that derides political elites in the ruling government as well as opposition parties. Twitter hashtags such as #PappuCII (dumb person at CII, or Confederation of Indian Industries) or #Feku (liar) reveal the *gaali*-fed framing of the opposition party leader and the prime minister, and the angst of the net-savvy youth against what is seen as inept political establishment. While such antiestablishment *gaali* does not always assume a progressive political position, the performative spread of *gaali* has nonetheless brought new political voices to the fore (Udupa, forthcoming). By their own account, *gaali*, as rancorous rabble-rousing, has helped them to thrust their voice into the public domain hitherto dominated by the state and organized commercial media. Contentions around corruption and governmental inefficiency and advocacy for net neutrality and free Internet are some of the vexing issues that have driven online middle-class engagement with mainstream politics in recent years. Online media have also enabled middle-class users to participate in global debates to challenge pervasive stereotypes about India and what they see as Western bias in portraying countries like India as backward Third World countries and to offer the counterimage of a rising global power. Numerous Twitter wars with abuse exchange to challenge global media representations of India testify to *gaali* culture as a struggle over meaning among middle-class online users of India.

The blurred arena of online comedy, insult, and abuse becomes vitriolic when the discussion centers on religious identities and ideas of nationalism gaining momentum on social media in recent times. Volunteers of right-wing religious nationalism advocating the cause of Hindu-first India figure prominently in abuse cultures. Scathing messages against minoritized Muslims and Christians are a common occurrence, and so are the acrimonious retorts and dodges among the avowed supporters of Muslim communities. This form of abuse culture centers on defining nationhood in terms of religious belonging. It has expanded in the historical context of religious majoritarianism, which considers non-Hindu religious groups as unfaithful citizens if they do not embrace Hinduism as a way of life.

For the right-wing Hindu nationalist volunteers on social media—shaped by not only top-down organizations, but by emergent networked publics—minority communities pose a challenge for national stability, but a bigger challenge is what they caustically dub as "pseudo secularism." India's secular politics, according to them, has betrayed the majority community by "appeasing" the minorities for electoral gains. On the other hand, the liberal position of the intelligentsia and a section of organized media is construed as hypocrisy of a privileged class who turn blind to the plight of majority Hindus, reflecting deeper tensions over the failures of secular liberal politics. On Twitter and Facebook, some of these contestations descend to abuse as they develop a distinct gendered character targeting dissenting voices through the masculinist logic of shaming. This signals the third variation of online abuse. Women journalists, academics, social activists, and cinema stars are often the target of this form of abuse. Social media users we met in Mumbai showed us abusive tweets and messages with verbal references to vagina, illicit sex, and prostitution (pimps, guttersnipes, *randi*/prostitute, bitch) in proses and sexist epithets. Sometimes these online *gaali* grow into a full-blown shaming punishment, articulating nationalism through the trope of regulating sexuality and what Irvine (1993) terms "evaluative talk." Online *gaali* as gendered

abuse was behind several high-profile cases of harassment, including, for instance, online attacks on Bollywood actor Shruti Seth and feminist activist Kavitha Krishnan in 2015, when they criticized the "selfie with daughter" online campaign initiated by the government, accusing the campaign of lacking the seriousness required to tackle grave issues of female infanticide and feticide (Seth, 2015).

Across multiple registers of online abuse, online users confront a tension between the "A" economy of anonymity (Auerbach, 2012), relying on an intentional disconnect between online and off-line selves and the economy of self-publicity with enumerative publicity measures. With abuse, online users find a way to be heard in the clutter of online traffic, and the allure of instant publicity props them to become even more cantankerous. At the same time, affordances of ambiguity ease the way to hurl invectives. Abuse is a paradigmatic practice in which online users hide and reveal themselves at the same time. This tension reflects the constitutive ambiguity of public cultures driven by a tension between publicity and containment (Mazzarella & Kaur, 2009).

The variety and ambiguity of abuse described, and not exhaustively by any means, reveal the irreducibility of online context that frames *gaali* as a particular Indian avatar of extreme culture. Hate speech not only fails to capture this range, but more seriously, collapses different user motivations and practices as a "single lump of fact" (Herzfeld, 2016). Extreme speech highlights precisely this ambiguity and variation in online vitriolic cultures, emphasizing the need to take into account different user groups and motivations as well as historical conditions of public speech cultures that lie behind invective exchange. A contextualized disaggregation of hate speech through the conceptual rubric of extreme speech is thus important not only to take account of the variety of online vitriol, but for turning a critical eye on the systematic use and management of online speech for political gains.

Hate Speech in Ethiopia

In contrast to India, where the government actively promotes Digital India initiatives and the online and mobile sectors are booming, Ethiopia remains one of the least digitally connected countries in the world. With Internet penetration rates hovering around 3% to 4% (Freedom House, 2015), Ethiopia's government has placed little emphasis on digital media as a driver for economic growth. It has censored websites with dissenting voices to its ideology of revolutionary democracy and developmental state as well as arrested bloggers, journalists, and other critical voices online. It seems as if the so-called Digital Ethiopia is underwritten by a curious paradox: If the Internet is considered so unimportant, then why is the government so afraid of it?

Ethiopia thus provides a unique case to understand vitriolic online speech and its implications for legal and political discourse. Similar to India, this term obfuscates more than it reveals. There are three specific challenges for understanding online hate speech debates in the Ethiopian context. First, there is no substantial ethnographic research on Ethiopia's online practices or the cultures of communication shaping them. The anthropological work that exists has focused mostly on the linguistic and cultural practices of rural communities or has looked at older media forms such as newspapers using the methods of interviews or content analysis (Skjerdal, 2016; Triulzi, 2006). Second, debates on hate speech have been historically linked to the shifting ethnic politics of post–civil war Ethiopia. Understanding online

speech in Ethiopia thus cannot be easily disentangled from these earlier and highly politicized and polarized debates, which have only now spilled over to the online sphere (see H. Ahmed, 2006; Barata, 2012; Tronvall, 2008. And, third, debates on hate speech in Ethiopia have been largely articulated through a developmental discourse where the onus of understanding has been on concerns over ethnic conflict and political stability. As a consequence, the research that does exist often has been framed through debates on conflict prevention with little emphasis on possible *emic* frames of online behavior. As Gagliardone, Kalemara et al. (2015) suggest, much of this discussion has been based on assumptions about "how they *ought* to work—assumptions typically defined by actors from the Global North" (p. 1). Given this background, a critical detour through the multiple cultural translations behind how hate speech has been reified in Ethiopia is needed to disentangle such complex political overdeterminations from local practices (see Asad, 1986).

Since coming to power after a 25-year bloody civil war, the ruling Ethiopian People's Revolutionary Democratic Party (EPRDF) has maintained a contentious relationship to online communication. This has been, in part, an attempt to balance freedom of expression with a desire to use media to further its own political goals. Gagliardone (2014) has extensively traced the genealogy of this contested relationship to the turbulent period that followed the overthrow of the communist dictatorship in 1991. He argues that the "original sin" of the government in the media sphere was that, while allowing free debates to proliferate following the end of the civil war, it refused to nonetheless engage with these debates because of its "belief that those writing for the private press were not part of the EPRDF's constituency . . . so there was little need to expend political capital either repressing or engaging them" (p. 285. Over time, however, this lack of engagement led to an increasingly polarized media environment where old grievances were amplified in the absence of real dialogue between the government and the opposition. So while countries such as Kenya were successfully experimenting with new digital technologies and online platforms, the Ethiopian government remained skeptical of their utility. Infrastructure development was slow, and the cost of connectivity kept its use out of the reach of the vast majority of Ethiopians, and especially the rural majority.

Yet despite the limited connectivity, the few Ethiopian online spaces that emerged became active, aided by a large diaspora who maintained a close interest in domestic politics. With diminishing space for free print media, people turned online, routinely printing news, commentaries, and political manifestos from these sites for broader distribution off-line. Moreover, it has been noted that during the heated debates around the contested 2005 elections, the language used in these online forums was highly charged with the rhetoric of hate speech, conjuring fears of ethnic violence from both the opposition and the government. Hate speech became a byword for the many social and political antagonisms in the country. Legesse (2012), for instance, has argued that there was "an unsettling resemblance between the hate propaganda used during the Rwandan genocide and the hate campaign surrounding the May 2005 elections in Ethiopia" (p. 360). He notes:

> In a country of ethnicized politics, talking about ethnic injustice is both inescapable and more likely to be punished as hate speech. Hence, what one observes under such circumstances is debate by proxy. Apparently neutral terms, such as "policy platform," "ideology," "economic policy," "land policy," and the like, are in fact loaded terms, code

> words for particular ethnicities. (p. 371)

More recently, these politicized debates on hate speech have also been linked to the 2009 Anti-Terrorism Proclamation, which empowers the Ethiopian government to punish speech that has the potential to "destabilize" the country with a risk of penalty of 15 years to life in prison or death. The state of emergency that followed civil unrest in 2016 has also resulted in the further criminalization of "negative" online and social media speech, described by the government as any kind of activity that "could create misunderstanding between people or unrest" (Human Rights Watch, 2016, para. 5.), conjuring once again the risk of metonymically substituting hate speech as a proxy for political dissent.

It is within this context that we held our first workshops in Ethiopia in 2014. Our aim was to bring together representatives from the Ethiopian government, opposition parties, civil society groups, academics, and journalists to discuss what many considered to be an escalating problem of hate speech in Ethiopian online spaces. In the first workshop, we thus wanted to establish an academic forum where research could be used to instigate a debate on these imagined dangers of online speech, and of hate speech in particular. We also presented the results of a pilot research, the first of its kind, where we had mapped out what we considered illustrative examples of vitriolic speech found in Ethiopian online discussions based on definitions used in international legal frameworks and especially the dangerous speech framework (Benesch, 2012; Sambuli et al., 2014). These findings resulted in a heated debate among the participants; while not agreeing on what the definition of hate speech was (or who was at fault for it), there was nonetheless an agreement among the participants about the urgent need to better understand hateful speech in the unique context of Ethiopian online communications. As one participant remarked:

> These are the horns on the head. This is hate speech. Fear. It is very obvious that what we are looking for here is the war for the mind. Dirty or clean, there is a war for the mind going on—pro and con government. What should we do?
>
> If the opposition succeeds in putting on big horns, people will be afraid. And vice versa with the government.
>
> Can we remove the horns?
>
> We are looking at conflict. In conflict we have two sides; in conflict we must make both sides "good"—two good groups to talk together. If one is entered as bad or good, there is no way forward.

As the workshop came to a close, there was optimism in the air. Instead of demonizing the other side of the political spectrum, perhaps grounded academic research could be used to create a dialogue between people who usually resorted to old political grievances where hate speech was concerned. Two months after the workshop, however, the Ethiopian government arrested more journalists and bloggers who were active online. The debates in Ethiopia were again as polarized as ever: The Ethiopian government was widely criticized by international organizations for violating freedom of speech and

human rights; the government, in turn, accused the journalists and bloggers for collaborating with foreign advocacy groups to stir up social instability and ethnic unrest through the use of social media.

We carried the lessons learned from the pilot to our second project on online hate speech in Ethiopia. This time, however, we adopted a new approach. Instead of focusing only on identifying illustrative examples of hate speech, we felt we needed to better understand the broader context of the social media debates in general. Our concern was that by selecting examples that were considered hate speech to begin with (based on some preexisting categories imposed from the outside), we risked creating an image of online conversations in Ethiopia as being harmful before we had done the research to assess whether this was the case. This posed two challenges for research. The first was how to methodologically place such examples of online vitriol within the broader context of online conversations in Ethiopia and its diaspora. The second was how to better understand the counterpart of hate speech—spaces of engagement—whereby users created communicative relationships and dialogue across political boundaries and divisions.

To achieve these two aims, we drew on the Ethiopian philosophical concept of *mechachal* to provide an alternative framework for approaching hate speech debates in the Ethiopian context. Girma (2012) defines this concept as follows:

> Mechachal is about one's own social sphere and the willingness to accommodate other social spheres that are different in a cultural or a religious sense . . . the essence of this concept is that it recognizes the pain involved in allowing plurality. And yet, it sees peaceful coexistence as something worth sacrificing for. (p. 181)

We then extensively debated our selected research methodologies in four collaborative workshops. That is, instead of automatically reverting back to the preexisting legal-normative frameworks for defining hate speech, we explicitly opened up this definition for a broader discussion among the workshop participants. Before each phase of the research, we thus validated the methodology, concepts, and our preliminary empirical results and in discussions with members from both sides of the political spectrum and the media. The end result was a broader definition of hate speech, which included multiple kinds of communicative relationships implicated in online speech in Ethiopia and not only the legal definition normally used to define it.

What was crucial about defining our object of study this way was that our empirical results surprised everybody, including ourselves. We found that online hate speech, at least according to the definition we had agreed upon in the workshops, was, in fact, marginal to Ethiopian online discussions. On the one hand, such vitriolic speech consisted mostly of conversations carried out by anonymous people with little influence or with only a few followers. On the other hand, the most relevant discussions online consisted of engaging types of conversations across boundaries. The fact that online speech had been framed as a problem for ethnic stability and as a catalyst for conflict in Ethiopia seemed to be without empirical grounding. It seemed as if the existing vitriolic online practices had been framed through powerful prearticulations shaped more by the prerogatives of domestic and international politics than by actual online practices behind them. Hate speech, from this perspective, seemed to act more as an "empty

signifier" (Laclau & Mouffe, 1985) onto which various domestic and global political fears and desires were projected than being an actual ontological object somewhere "out there." Our research thus concluded that, while such examples of vitriolic speech can of course be found across online conversations in Ethiopia (as they can be found anywhere else) and should be taken seriously because they are reflective of the broader ethnic and political tensions in the country that we have seen escalate over the past year, social media in Ethiopia seems to be nonetheless also "emerging as a space where different forms of tolerance and acceptance are being displayed and new forms of engagement can be experimented with" (Gagliardone et al., 2016, p. 10).

With this relationship between hate speech and politics in mind, an important set of questions is raised for understanding the concept of extreme speech in the Ethiopian context: Given the powerful preexisting articulations, how can we best identify online practices in Ethiopia that would be comparable to, for instance, *gaali* in India? The answer to this question can only come from more ethnographic work on online and social media communities in Ethiopia and its diaspora. Concepts such as *mechachal* that we have started with have provided preliminary ways of intervening in existing debates. There is a need for more ethnographic appreciation of online speech based on categories defined from within that would shift focus to the everyday practices behind how people use, and often subvert, online communication for their own purposes.

Indeed, as the first theoretical move toward this aim, we could begin by highlighting existing Ethiopian traditions that can provide critical alternatives to the existing and politicized hate speech debates. For instance, one of the best known traditions in Ethiopia consists of the various liturgical and literary practices loosely called the *Sen-ena-Werq* (wax and gold).[4] This tradition, derived from the highland Amhara and the Ethiopian Orthodox Church, consists of a philosophy of language/meaning that is based on complex double meanings, wordplay, and the use of metaphor. While the merits of using this concept for political and social analysis has been highly contested among researchers (see Levine, 1965; Messay, 1999), the wax and gold tradition, by foregrounding ambiguity and subterfuge, could critically sidestep a more deterministic understanding of communication by bringing it closer to a more poststructuralist understanding of communication as a form of play and creative appropriation of existing norms against power (see Baudrillard, 1983; Bezabeh, 2014). Girma (2011) writes that "the wax and gold's affinity with dualism seems to have served an unintended purpose—the ambiguity surrounding it, at times, seems to have provided a space in which to criticize people who otherwise are hard to reprimand" (p. 176). Levine (1965) has similarly noted that this tradition of communication has historically been used as a "means to insult somebody in a socially acceptable manner," as a technique for "defending the sphere of privacy against excessive intrusion," and as a "medium to criticize authority" (p. 9) in a society where violent conflict and excesses of authority have always been dominant.

Similarly, research on other Ethiopian traditions of communications such as the *Gadaa* among the Oromo, the *Kassow songs* among the Afar, or the *Gubo* among the Somali communities can also potentially provide repertoires of new concepts to initiate such critical anthropological research on online

[4] There are many variations on how to spell this term. We have chosen *Sen-ena-Werg* as translated by Girma (2012).

communication in Ethiopia. Whether these traditions are suitable for understanding contemporary online practices in Ethiopia and especially its hybrid diasporic online spaces remains too early to speculate (see Skjerdal, 2009). What our discussion of the concept of extreme speech in the Ethiopian context, however, suggests is that defining such speech without first closely understanding its illocutionary contexts (Butler, 1997) can obviate the subtle nuances of communication behind overarching categories such as hate speech if the definition is imposed from the outside.

Conclusion

As is evident from the cases of India and Ethiopia, the textured nature of online abuse and invective language belies the presuppositions of umbrella concepts such as hate speech as well as the celebratory discourse of online subversion. Whether the case is of the *gaali* tradition in India or distinct linguistic traditions in Ethiopia, the diverse online practices defy easy categorizations that could be mapped onto a bipolar field of acceptable and unacceptable speech. This ambiguity underlies the regulatory dilemma between free speech and hate speech. The dilemma has produced a veritable mass of political technologies and legal regulation where both camps have assiduously chased the principles that are at odds with one another. Anthropology, we suggest, might provide one way through this impasse. As Michael Herzfeld (2016) summarizes, the problem of culture and relativism central to anthropological practice brings back the ethical responsibility of evaluation. "Restrained relativism," as he calls it, restores the key anthropological principle of "context"; "and context, in turn, takes the ethical back down from the level of abstraction that would countenance any and every cultural principle on the grounds that some group endorses it" (p. 2). Yet, at the same time, this evaluation cannot lead to a reductionist "audit culture" or "ranking culture" based on prior categorical templates (Herzfeld, 2016). At worst, this reductionism can lead to evaluation and judgment as rhetorical and political practice, flattening contexts, histories, and practices on an enumerative scale of metrics and data. A clear offshoot of this audit culture is the growing prominence of the "risk discourse," where, in the name of ethical evaluation, public interest, and safety, an expanding machinery of surveillance and censorship has been created to mitigate the imagined dangers of unruly online conduct and with little evidence about its real dangers. Digital policy should be careful, at the least, of such regulatory excesses guised under the templates of hate speech while being attentive to various effects, including reactionary and violent consequences, of online speech.

Our concept of extreme speech has thus been an attempt to move the debate beyond a normative understanding of vitriolic online speech practices as hate speech. In its place, in line with comparative practice, we propose a two-pronged typology for researching online vitriol in different parts of the world. The first move requires us to critically trace the genealogy through which existing hate speech discourse has been superimposed onto a multiplicity of digital cultures in different regions. The second move consists of developing a more situated understanding of the cultures of communication and online practices that have been obfuscated by the overarching category of hate speech. In place of universalizing normative frameworks, we suggest an anthropological approach that would dissect what people do with online media—and, moreover, how they themselves understand and comment on the significance of their own actions and its imagined intentions and causes. In other words, extreme speech calls for a critical typology based on the contexts of digital use, distinct histories and public cultures behind different online

speech forms, variations in user motivations, and the complex politics involved in labeling certain kinds of speech as one thing or another (Udupa, 2016). But it also includes, perhaps more fundamentally and radically, opening up digital research to some of the more *emic* categories through which the complex use of language and, ultimately, our understanding of communication operate based on different philosophical registers outside the West. As our discussion of India and Ethiopia has hinted at, even media theory—with its metaphysics firmly grounded in Western philosophical traditions of communication—might not, in the end, be universally applicable for understanding online practices in other parts of the world. Our proposed concept of extreme speech, thus, hopes to bring to the fore contextual differences as the majority of the world's populations becomes connected to the Internet and begins to communicate in ways we may have not even anticipated.

References

Ahmed, A. A. (2009). Specters of Macaulay: Blasphemy, the Indian penal code and Pakistan's postcolonial predicament. In R. Kaur & W. Mazzarella (Eds.), *Censorship in South Asia: Cultural regulation from sedition to seduction* (pp. 172–205). Bloomington: Indiana University Press.

Ahmed, H. (2006). Coexistence and/or confrontation? Toward a reappraisal of Christian-Muslim encounter in contemporary Ethiopia. *Journal of Religion in Africa*, *36*(1), 4–22.

Amoore, L., & Goede, M. D. (2008). *Risk and the war on terror*. New York, NY: Routledge.

Asad, T. (1986). The concept of cultural translation in British social anthropology. In J. Clifford & G. E. Marcus (Eds.), *Writing culture: The poetics and politics of ethnography* (pp. 141–164). Los Angeles: University of California Press.

Auerbach, D. (2012). *Anonymity as culture: Treatise*. Retrieved from https://www.canopycanopycanopy.com/contents/anonymity_as_culture__treatise

Barata, D. D. (2012). Minority rights, culture, and Ethiopia's "third way" to governance. *African Studies Review*, *55*, 61–80.

Baudrillard, J. (1983). *Simulations*. New York, NY: Semiotext(e).

Beck, U. (2000). *What is globalization?* Malden, MA: Polity.

Benesch, S. (2012). Dangerous speech: A proposal to prevent group violence. Retrieved from http://www.worldpolicy.org/sites/default/files/Dangerous%20Speech%20Guidelines%20Benesch%20January%202012.pdf

Bezabeh, S. A. (2014). Living across digital landscapes: Muslims, Orthodox Christians and an Indian guru in Ethiopia. In R. I. J. Hackett & B. F. Soares (Eds.), *New media and religious transformations in Africa* (pp. 266–283). Bloomington: Indiana University Press.

Brubaker, R., & Cooper, F. (2000). Beyond "identity." *Theory and Society*, *29*(1), 1–47.

Butler, J. (1997). *Excitable speech: A politics of the performative*. New York, NY: Routledge.

Buyse, A. (2014). Words of violence: "Fear speech," or how violent conflict escalation relates to the freedom of expression. *Human Rights Quarterly*, *36*(4), 779–797.

Couldry, N. (2010). Theorizing media as practice. In B. Brauchler & J. Postill (Eds.), *Theorizing media and practice: Anthropology of media* (pp. 35–55). Oxford, UK: Berghahn Books.

Facebook. (2017). What does Facebook consider to be hate speech? Retrieved from https://www.facebook.com/help/135402139904490/

Freedom House. (2015, October). *Freedom on the net 2015: Privatizing censorship, eroding privacy*. Washington, DC: Author. Retrieved from https://freedomhouse.org/sites/default/files/FOTN%202015%20Full%20Report.pdf

Fuchs, C. (2015). *Culture and economy in the age of social media*. New York, NY: Routledge.

Gagliardone, I. (2014). New media and the developmental state in Ethiopia. *African Affairs*, *113*(451), 279–299.

Gagliardone, I., Kalemara, A., Kogen, L., Nalwoga, L., Stremlau, N., & Wairagala, W. (2015). In search of local knowledge on ICTs in Africa. *Stability: International Journal of Security and Development*, *4*(1), 1–15.

Gagliardone, I., Pohjonen, M., Zerai, I., Beyene, Z., Aynekulu, G., Stremlau, N.,…Gebrewolde, T.M. (2016). *MECHACHAL: Online debates and elections in Ethiopia. From hate speech to engagement in social media*. University of Oxford Programme in Comparative Media Law and Policy. Retrieved from https://www.academia.edu/25747549/Mechachal_Online_Debates_and_Elections_in_Ethiopia._Final_Report_From_hate_speech_to_engagement_in_social_media_Full_Report_

Geertz, C. (1973). *The interpretation of cultures: Selected essays*. New York, NY: Basic Books.

Girma, M. (2011). Whose meaning? The wax and gold tradition as a philosophical foundation for an Ethiopian hermeneutic. *Sophia*, *50*(1), 175–187.

Girma, M. (2012). *Understanding religion and social change in Ethiopia: Toward a hermeneutic of covenant*. New York, NY: Palgrave Macmillan.

Hardaker, C. (2010). Trolling in asynchronous computer-mediated communication: From user discussions to academic definitions. *Journal of Politeness Research*, *6*(2), 215–242.

Hate speech: SC fixes Subramanian Swamy's plea for final hearing. (2017, January 9). *The Indian Express*. Retrieved from http://indianexpress.com/article/india/hate-speech-sc-fixes-subramanian-swamys-plea-for-final-hearing-4466924/

Herzfeld, M. (2016, April 8). *Comparing values: Ethics, audit culture, and the menace of reductionism—The Value of Comparison symposium*. Goettingen, Germany: Max Planck Institute for the Study of Religious and Ethnic Diversity.

Human Rights Watch. (2016, October 30). *Legal analysis of Ethiopia's state of emergency*. New York, NY: Author. Retrieved from https://www.hrw.org/news/2016/10/30/legal-analysis-ethiopias-state-emergency

Irvine, J. T. (1993). Insult and responsibility: Verbal abuse in a Wolof village. In J. H. Hill & J. T. Irvine (Eds.), *Responsibility and evidence in oral discourse* (pp. 105–134). Cambridge, UK: Cambridge University Press.

Laclau, E., & Mouffe, C. (1985). *Hegemony and socialist strategy*. London, UK: Routledge.

Legesse, Y. M. (2012). Shielding marginalized groups from verbal assaults without abusing hate speech laws. In M. Herz & P. Molnar (Eds.), *The content and context of hate speech* (pp. 352–377). Cambridge, UK: Cambridge University Press.

Levine, D. (1965). *Wax and gold: Tradition and innovation in Ethiopian culture*. Chicago, IL: University of Chicago Press.

Lovink, G. (2013). Hermes on the Hudson: Notes on media theory after Snowden. *E-Flux*, 54. Retrieved from http://www.e-flux.com/journal/hermes-on-the-hudson-notes-on-media-theory-after-snowden/

Marwick, A., & boyd, d. (2011). *The drama! Teen conflict, gossip and bullying in networked publics*. Oxford, UK: Oxford Internet Institute.

Mazzarella, W., & Kaur, R. (2009). Between sedition and seduction: Thinking censorship in South Asia. In R. Kaur & W. Mazzarella (Eds.), *Censorship in South Asia: Cultural regulation from sedition to seduction* (pp. 1–28). Bloomington: Indiana University Press.

Messay, K. (1999). *Survival and modernization: Ethiopia's enigmatic present: A philosophical discourse*. Asmara, Eritrea: Red Sea Press.

Morozov, E. (2012). *The net delusion: The dark side of Internet freedom*. New York, NY: Public Affairs.

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). New Delhi, India: SAGE Publications.

Nockleby, J. T. (2000). Hate speech. In K. L. Leonard & W. Levy (Eds.), *Encyclopedia of the American constitution* (pp. 1277–1279). New York, NY: Macmillan.

Parliament of India. (2015, December 7). *Action taken by the government on the recommendations/observations contained in the 176th report on the functioning of Delhi police* (Report No. 189). New Delhi, India: Rajya Sabha Secretariat. Retrieved from http://164.100.47.5/newcommittee/reports/EnglishCommittees/Committee%20on%20Home%20Affairs/189.pdf

Rajagopal, A. (2001). *Politics after television: Hindu nationalism and the reshaping of the Indian public*. Cambridge, UK: Cambridge University Press.

Rao, K. (2016, October 24). EIC outrage: Free speech [Video]. East India Comedy. Retrieved from https://www.youtube.com/watch?v=3zxNCuH5R2Y

Rowbottom, J. (2012). To rant, vent and converse: Protecting low level digital speech. *Cambridge Law Journal*, *71*(2), 355–383.

Samaratunge, S., & Hattotuwa S. (2014). *Liking violence: A study of hate speech on Facebook in Sri Lanka*. Colombo, Sri Lanka: Centre for Political Alternatives.

Sambuli, N., Morara, F., & Mahihu, C. (2014). *Monitoring online dangerous speech in Kenya: January–November 2013* (Umati Report). Retrieved from http://ihub.co.ke/blog/wp-content/uploads/2014/06/2013-report-1.pdf

Seth, S. (2015, July 2). A little note to India [Blog post]. Retrieved from http://www.twitlonger.com/show/n_1smtdi6

Skjerdal, T. (2009). A critical look at the digital diaspora: Perspectives from Ethiopia. In K. S. Orgeret & H. Rønning (Eds.), *The power of communication: Changes and challenges in African media* (pp. 311–348). Oslo, Norway: Unipub.

Skjerdal, T. (2016). Online journalism under pressure: An Ethiopian account. In H. M. Mabweazara, O. F. Mudhai, & J. Whittaker (Eds.), *Online journalism in Africa: Trends, practices and emerging cultures* (pp. 89–103). London, UK: Routledge.

Triulzi, A. (2006). When orality turns to writing: Two documents from Wallaga, Ethiopia. *Journal of African Cultural Studies*, *18*(1), 43–55.

Tronvoll, K. (2008). Human rights violations in federal Ethiopia: When ethnic identity is a political stigma. *International Journal on Minority and Group Rights*, *15*, 49–79.

Udupa, S. (2015). *Making news in global India: Media, publics, politics*. Cambridge, UK: Cambridge University Press.

Udupa, S. (2016, March 14). Middle class on steroids: Digital media politics in urban India. *India in Transition*. Center for the Advanced Study of India, University of Pennsylvania. Retrieved from https://casi.sas.upenn.edu/iit/sudupa

Udupa, S. (forthcoming). *Gaali* cultures: The politics of abusive exchange on social media. *New Media & Society*.

Van der Veer, P. (2001). *Imperial encounters: Religion and modernity in India and Britain*. Princeton, NJ: Princeton University Press.

Van der Veer, P. (2016). *The value of comparison*. Durham, NC: Duke University Press.

Waldron, J. (2012). *The harm in hate speech*. Cambridge, MA: Harvard University Press.

Wang, T. (2013, August 3). Our commitment [Blog post]. Retrieved from https://blog.twitter.com/en-gb/2013/our-commitment

Warner, W., & Hirschberg, J. (2012). Detecting hate speech on the World Wide Web. In *Proceedings of the 2012 workshop on language in social media* (pp. 19–26). Montreal, Canada: Association for Computational Linguistics.

YouTube. (2017). Hate speech. Retrieved from https://support.google.com/youtube/answer/2801939