Introduction

Lived Encryptions

WhatsApp, Disinformation, and Extreme Speech

SAHANA UDUPA AND HERMAN WASSERMAN

In the months leading up to Chile's historic 2022 referendum on its proposed new constitution, WhatsApp groups in the country saw a significant surge in conspiracy theories about election fraud, as misinformation messages that were predominant within right-leaning groups in the beginning soon spread and "contaminated" left-leaning groups that were in favor of the proposal. During the fiercely contested presidential elections in Brazil in 2018 and 2022, several women in the favelas received messages in the WhatsApp church group, hailing right-wing conservative leader Jair Bolsonaro as the "man of God." There was little that could be disputed, in their mind, that "God, homeland and family" are deeply connected, and amid waves of WhatsApp messages that echoed such sentiments, they switched their loyalties from progressive parties to Bolsonaro's conservatism.

Anti-Black far-right ideologies are common within WhatsApp groups of "Operation Dudula," an anti-immigration group in South Africa, although a majority of its users are Black South Africans. Contrary to assumptions that hateful narratives ride on brazen falsehoods, members of this group spend much time to offer "accurate" information as a rhetorical ploy to ensure their groups are free from "inauthentic" and "criminal" groups who could spoil their "brand image" and upset their activities to prevent immigration from neighboring African countries into South Africa.

At traffic checkpoints in the Anglophone regions in Cameroon, amidst an escalating conflict between separatist and state forces, it is a routine practice for government authorities to stop passengers and demand to see not only their national identity cards but also their smartphones to

manually check if they carry any "incendiary" messages. WhatsApp's encryption is overruled in such swift everyday acts of surveillance.

On WhatsApp groups in India, Hindu nationalists have transformed encryption from a technical feature of privacy to one that can foster obligatory and affective chambers for ideological talk. Disinformation proliferates within WhatsApp groups modeled as kin-like groups. Similarly, in South Africa and Kenya, convivial bonds within communities breed disinformation on WhatsApp groups since any act of correcting dubious messages passed on by known members of family or church members is viewed as impolite or disrespectful of one's elders.

As these examples from different contributions in this volume illustrate, ironies, contradictions, and thick social norms and community affect suffuse WhatsApp discourse, while their ramifications have contributed to invasive state surveillance and some of the worst human tragedies, most significantly in the diverse contexts of the Global South, which are marked by historical exploitation and injustices as well as persistent socioeconomic inequality.

Disinformation and vitriolic expressions on WhatsApp have received extensive media and public policy attention, alongside academic scholarship that has concurrently drawn attention to the broader and complex ecosystems of hate and disinformation in the digital age. Recent academic scholarship has taken note of the risks of internet communication for democratic systems, as regressive regimes around the world have weaponized online discourse for partisan gains during elections, to undercut domestic dissent or power up geopolitical contestations against "rival" nation states through targeted disinformation campaigns (Bayer and Bárd 2020; George 2016; Graan, Hodges, and Stalcup 2020; Krafft and Donovan 2020; Lee 2019). Acknowledging the role of digital networks in inspiring social movements to hold the power to account and question entrenched hierarchies, studies have simultaneously highlighted the worrying developments around vitriolic exchange, inauthentic content, and "antisocial commenting" that are breeding on affordable communication and the circumvention of legacy gatekeepers that digital media infrastructures have enabled (Shmargad et al. 2024, 220).

The extent of disinformation and extreme speech prevalent in online exchanges is notoriously slippery for quantification, especially due to a lack of access to social media company data, including comprehensive

transparency reports, and various barriers that companies have raised for API-based data gathering and auditing (Freelon 2018). Facebook reported that "between January 2021 and March 2021, there was a 0.05 percent to 0.06 percent prevalence of hate speech, showing a slight decrease compared to their two previous reports" (Bright et al. 2021, 6), although processes of drawing such metrics remain inaccessible to researchers. It is indeed not common for social media companies to publish "prevalence metrics," and "evidence about the prevalence of hate speech on social media platforms remains incomplete, partly due to a lack of transparency and data access on the part of platforms" (2021, 6).

While no consensus in academic scholarship about quantitative indicators of "prevalence" has followed as a result, academic studies, policy reports, and regulatory measures have highlighted the significance of the issue both in terms of public perceptions and emerging patterns of public discourses. According to the World Economic Forum *Global Risks Report*, misinformation and disinformation "was perceived as a moderately severe risk" by its respondents (World Economic Forum 2023, 24), while the Ipsos and UNESCO report (2023, 21–23) identified social media as a "top source of information in every country" it surveyed, finding that "two thirds [of respondents] often encounter hate speech online." Users in the Global South regions themselves experience the impact of disinformation on their daily lives and political participation as a major problem. Research in several sub-Saharan African countries, for instance, has shown that perceived exposure to disinformation is high and is linked to low levels of trust in social and national media (Wasserman and Madrid-Morales 2019).

Importantly, studies have shown that disinformation, misinformation, and extreme speech feature prominently in political discourses during major developments or events such as elections, social movements, and referendums within diverse national, local, and translocal contexts as well as linked to global strategic interests of "foreign actors" (Bradshaw and Howard 2018). In the Anglophone crisis in Cameroon, Schumann (chapter 5, this volume) notes that online posts are made every day and some incidents (fifty-two at the time of publication) have been verified by reports that are archived in the Cameroon Database of Atrocities hosted by the University of Toronto (Borealis 2024). Nkululeko and Gagliardone

(this volume) cite the *Disinformation in Africa 2024* report published by the Center for Strategic Studies, which has found 189 documented disinformation campaigns in Africa sponsored by foreign governments including China and Russia, noting that such campaigns have quadrupled since 2022. Incidents of misinformation, disinformation, and deep fakes, including the use of artificial intelligence, have been documented around the world (Kertysova 2018). Subsequently, the number of publications on hate speech and related phenomena has seen an exponential growth between 1992 and 2018 (Tontodimamma et al. 2021, 163) and it continues to rise (Walther and Rice 2024).

In the complex mix of factors that shape extreme speech ecosystems, encrypted instant messaging services such as WhatsApp, Telegram, and Signal constitute a unique constellation. Such messaging services lack some of the core features of networked communication that typify social media, most prominently the ability for "participants [to] have uniquely identifiable profiles that consist of user-supplied content, content provided by other users, and/or system-level data . . . [and to] . . . publicly articulate connections that can be viewed and traversed by others" (Ellison and boyd 2013, 533). Content shared in one WhatsApp group, for instance, cannot easily reach other groups, and users on the platform do not have the affordances to present profiles based on metadata such as likes and followers and algorithmically curated metrics; they also cannot publicly maneuver their lists of contacts. In the messaging app ecosystem, influencers are likely to be those who contribute the most or whose posts are shared and liked most frequently, but such metrics are not easily available to gauge nor are they appended to the profile. WhatsApp, like other instant messaging services, is also distinct for its chronological message display and the conspicuous lack of algorithmic feed that characterizes quasi-public social media platforms.

However, alongside the basic feature to "consume, produce and/or interact with streams of user-generated content provided by their connections" (Ellison and boyd 2013, 533), encrypted messaging is integrating social media-type functionalities such as group messaging, bulk forwards, user reactions, channels, and group lists, thereby transitioning from a strictly interpersonal form of communication to a social media-like platform. Importantly, in contrast to social media channels that promise publicity and visible, and even spectacular, disruptions of mainstream

media and political discourses, encrypted messaging often flows below the ground and end-to-end, often slipping out of direct regulatory reach and academic scrutiny. The very encrypted nature that makes messaging services attractive as a secure communication infrastructure also complicates access for research and content moderation, thus raising serious methodological and regulatory challenges.

This volume takes this vastly popular and important form of internet-enabled communication for closer examination, to develop a global critical inquiry into entanglements between encryption and extreme speech. It approaches the problem with an empirical focus on WhatsApp—an end-to-end encrypted, cross-platform messaging service owned by Meta (which purchased the app in 2014)—which has emerged as a central communication tool for a large number of people, with more than two billion users and one hundred billion daily messaging the world over (Ceci 2022). While messaging services such as Signal, Telegram, and WhatsApp have comparable functionalities and pose similar regulatory challenges to "securely" screen encrypted messages (Guest 2023), the popularity, reach, and specific purposes to which such messaging apps are put to use depend on the broader media and political systems they are embedded within (Rogers 2020; Semenzin and Bainotti 2020).

This volume explores WhatsApp as a critical window to encrypted messaging as a globally prevalent form of communication, particularly influential as a communication channel and platform for social and political mobilizations in the Global South. The analytical and methodological lessons derived from studying this messaging app, which has the largest user volume among such apps globally, can provide useful perspectives to study other messaging services as well as digitally mediated disinformation and extreme speech more broadly. WhatsApp's popularity has been linked to low internet connectivity and high data costs in the Global South contexts, but its uptake in different parts of the world, including in the Global North regions, awaits systematic research on user practices, infrastructural conditions, and political deployments around encrypted messaging, and how such features have uniquely inflected disinformation and extreme speech environments.

With a set of cross-disciplinary studies on a range of national and transnational contexts, and focusing especially on the Global South, we

address this gap and propose the concept of "lived encryptions." Turning to practices where WhatsApp intersects with vastly complex social and political fields and the lived worlds of users, "lived encryptions" stresses that encryption as a technological feature cannot be taken at its face value or as a central piece of the affordance as it is experienced; rather, it embeds different, often contradictory, social and political formations and interactions. This is evidenced, for instance, in the way the promised confidentiality of encrypted messaging is upturned completely when surveilling states seize the phones from suspected dissenters to download the data or how seemingly closed group communication is channelized to "broadcast" top-down political messages. In the Global South contexts, the emergence of "broadcast" WhatsApp groups testifies to novel ways of creating conditions of virality where unfamiliar senders invert the very logic of end-to-end encryption as a privacy- inducing feature and transform it into a subsidiary of community conversation that can render political messages as socially significant. As such, encryption does not unequivocally pave the way for feelings of safety and security; within conflict and authoritarian contexts, it triggers tense appraisals around safe and unsafe spaces. The conceptualization of "lived encryptions" foregrounds such tensions, accounting for irreducible user cultures and localized innovations for political propaganda in theorizations of digital communication, disinformation, and vitriol. It also highlights methodological difficulties of tracking them, and the value of ethnographic research in navigating intimate networks of messaging that are hard to access by other methodological means, as well as in contextualizing the varied contradictions of encryption.

In the rest of this introduction chapter, we will outline key observations around WhatsApp, disinformation, and vitriol in current scholarship, lay out the theoretical stakes of encryption and develop the framework of "lived encryptions." Throughout this discussion, we will closely converse with the contributions in this volume to delineate the contours of what we call "lived encryptions." The final section provides a description of the five sections featured in this volume—politics of divisive messaging, safe/unsafe spaces, infrastructure, method, and policy—and how they advance distinct yet interconnected lines of inquiry around WhatsApp.

Hate and Disinformation on WhatsApp

Scholars have documented the vast popularity of instant messaging services, describing WhatsApp, for instance, as a "technology of life" in the Global South (Cruz and Harindranath 2020). Studies on Africa have stated that "WhatsApp is the internet and vice versa" (Mare and Munoriyarwa, this volume), and scholars on South America have noted how WhatsApp, affectionately called by the name "ZapZap," is an entrenched everyday communication infrastructure in Brazilian favelas (Parreiras, this volume) and a primary communication tool in Chile (Santos, Ortiz Fuentes, and dos Santos, this volume). The vast popularity of WhatsApp has emerged from path dependency and low data usage, both tied to the political economy of digital media expansion in the South when companies like Meta (formerly Facebook) adopted a predatory path to establish new user bases with aggressive acquisition strategies and programs such as Free Basics (free internet in exchange for default provision of company's social media platforms and by limiting the access to other platforms). As a result, mobile operators in several countries in the South offer zero-rated data plans for WhatsApp and Facebook access, while the company has also introduced more features to offer commercial services to "business users" and digital payments in some markets (Cruz and Harindranath 2020).

African grassroots movements, civil society organizations, and concerned citizens have "leveraged the platform to raise awareness, advocate for issues, and mobilize support," as Olaniran points out in his contribution to this volume, citing social and political movements in Nigeria such as the #BringBackOurGirls campaign and the #EndSARS protest. The possibility to share news articles, videos, and personal accounts on WhatsApp, he points out, has enabled users to create a "network of engaged citizens who could drive change and influence public opinion." Wasserman and Madrid-Morales in this volume similarly discuss how WhatsApp has offered multiple avenues for communication among different communities in South Africa.

In contexts where other social media platforms are banned or restricted, for instance the Russian extremism law which does not allow Facebook, Instagram, and Twitter to operate in the country, users have

moved to messaging services not only to hold private conversations but increasingly to access news and public information (Sauer 2022). Encrypted channels are also popular for "political talk" in Europe, especially among users who are "reluctant to talk about politics in public," as evidenced in parts of former East Germany (Valeriani and Vaccari 2017). The *Digital News Report* by the Reuters Institute for the Study of Journalism at the University of Oxford has provided evidence that more users are holding political discussions through WhatsApp, often considered as "private" discussions on politics in contrast to perceptions of public engagements (Newman et al. 2023).

Simultaneously, WhatsApp's role in electoral politics and ecosystems of spurious content has expanded (Garimella and Eckles 2020). As Cheeseman et al. (2020, 145) note, "In the space of just a year, countries as otherwise diverse as Brazil, India, and Nigeria were said—with varying degrees of accuracy—to have witnessed their first 'WhatsApp election,' with the dissemination of rumors, conjecture, and lies allegedly undermining the democratic process itself." Rossini et al. (2021, 2434) have found "clear evidence that WhatsApp has been successfully used to spread false and misleading information during elections in Brazil as well as in India and Indonesia." Social media and messaging platforms do not remain hermetically sealed, however. Cross-fertilization between platforms, such as amplification of WhatsApp messages on Twitter and Facebook, has been documented in contexts such as Nigeria and India. In addition, recent developments of "deplatforming" hate influencers from Facebook and Twitter have spurred a wave of platform migration, as far-right actors in Europe and North America have turned to encrypted messaging alongside smaller platforms to build "resilience" and sustain group mobilization (Rogers 2020).

Disinformation and hate speech scholarship has largely advanced inquiries around instant messaging services within a broader critical framework which posits that the rules, protocols, conventions, and default designs of social media platforms shape user interfaces to enable and constrain forms of communication (Manovich 2001). Aside from the central feature of end-to-end encryption, topical studies have foregrounded the significance of closed communication architecture in messaging services as opposed to a timeline-based news feed and the absence of algorithmic sorting of content common in quasi-public social

media platforms such as Twitter or Facebook. In addition, instant messaging services offer ways to imbue messages with ephemerality with the functionality of "disappearing messages." In contrast to radical user anonymity and subcultural semiotics of niche small platforms such as 4Chan (Auerbach 2012; Knuttila 2011), sources of WhatsApp messaging are both known and unknown, as it spreads within closed communication groups and among members whose telephone numbers are visible, yet elusive when groups expand beyond direct contacts. In addition, the possibility to share texts, voice notes, still and moving images, and weblinks creates a multimodal and flexible conversational environment. Empirical evidence on partisan and inflammatory content flowing through WhatsApp has occasioned the argument that platform features, including affordances of forwards, group chats, and group calls, have prominently contributed to the ease and amplification of disinformation, conspiracy theories, and vitriolic exchanges (Binder, Ueberwasser, and Stark 2020; Evangelista and Bruno 2019; Johns and Cheong 2021; Nizaruddin 2021; Recuero, Soares, and Vinhas 2021; Resende et al. 2019b; Rossini et al. 2021; Soares et al. 2021; Williams et al. 2022).

Qualifying platform-based analysis, anthropologists have adopted a media practice framework, asking what people do with media and how complex mediations of lived worlds cluster around, draw upon, and reshape technological possibilities. Ethnographic and interdisciplinary studies on hateful speech and disinformation in the Global South have especially drawn attention to the political use and party deployments of WhatsApp (Pinheiro-Machado and Vargas-Maia 2023; Wasserman and Madrid-Morales 2022). As Olaniran discusses in his chapter in this volume, Nigerian politicians have exploited WhatsApp's affordance of connecting members of communities by creating WhatsApp groups that serve as hubs for their supporters and volunteers. He shows how, by fostering personal connections, politicians aim to cultivate loyalty, mobilize their base, and disseminate their political messages more effectively. Politicians also tap into existing social, religious, or community networks on the platform to amplify their messages.

Such political deployments of WhatsApp are also strikingly illustrated by the prominent role of Bharatiya Janata Party (BJP), the right-wing nationalist party in India, in engaging cross-media manipulation through organized circulation of "trend alerts" on WhatsApp groups (Jakesch

et al. 2021), spurring grave incidents of mob lynching (Vasudeva and Barkdull 2020). Similarly, WhatsApp's deployment in right-wing conservative leader Jair Bolsonaro's campaigns in Brazil has been widely documented (Machado et al. 2019; Parreiras, this volume).

Lived Encryptions

Picking up insightful threads from this scholarship, we center the significance of encryption as a technological infrastructure, social condition, and regulatory target. Our emphasis on encryption stems from the core feature of WhatsApp, which distinguishes it not only from other "open" social media platforms but also the direct messaging function available on such platforms. Our point of departure is the framework of "extreme speech," which refers to speech acts (text, audio, video, multimodal) that stretch the boundaries of legitimate speech along the twin axes of truth/falsity and civility/incivility (Udupa 2018b; Udupa and Pohjonen 2019). Distinct from the universal conception of "hate speech" and the risks of its regulatory misuse, the extreme speech framework stresses on ethnographic sensibility to cultural variation, historical awareness, and ambiguity of vitriol in assessing the nature and implications of contentious content. Of primary concern is the ways in which users draw meanings and create networks of distribution of extreme narratives, and sociocultural and historical factors that coalesce to shape the trajectories and consequences of such narratives. Methodologically, it calls for multiorder analysis linking platform features, user practices, and historical and political contexts surrounding digital messaging.

Guided by the extreme speech framework, we center the analytical value of encryption and consider encryption not as a determining technology feature but one that is suffused with multiple articulations and ridden with contradictions, which we capture as "lived encryptions."

In its unfolding, lived encryptions embed intimacy in articulating with the closed messaging architecture of messaging services, thereby enabling a sense of community that emerges within chat communication. For sure, this sense of community is neither exclusive to nor bounded by chat architecture but the very boundaries that closed chat architecture afford can ease the way to get a sense of community. At a bare minimum, this community is a communicative formation; i.e., one

belongs to the community because one reads, shares, likes, and posts messages with others and develops a degree of intimacy because of staying communicative within bounded loops of WhatsApp groups.

Closely tied to the conditions of intimacy is the possibility of trust and a shared feeling of "tight-knit networks" (Belinskaya and Rodriguez-Amat, this volume) and greater user control over contacts, if not always over content. Conditions of intimacy and trust are neither determined by nor reducible to technology design, but they are vitally linked to social narratives and shared perceptions. In Turkey, anthropologist Erkan Saka (this volume) says that "WhatsApp is assumed to be more social, more familial than other . . . services." In many cases, existing social groups duplicate as WhatsApp groups, transferring trust and intimacy along the way. In the favelas of São Paulo, WhatsApp users who spoke to anthropologist Carolina Parreiras insisted that Workers Party (PT) leader and former Brazilian president Dilma Vana Rousseff's character was questionable because they had received messages that portrayed her in a negative light from friends and acquaintances "she trusted" (this volume).

Conditions of intimacy and trust shaped by encryption-enabled closed architecture have encouraged political actors to enlist WhatsApp groups to create and disseminate extreme content in intrusive forms, enabling what is described as "deep extreme speech"—forms of discourse in which exclusionary content comes comingled with good morning greetings and pleasant messages, "simulating the lived rhythm of the social" (Udupa, this volume). Such socially sanctified messages become widespread when political parties link WhatsApp groups through volunteer and party mediators, drawing them into networks that connect the desired narrative across other social media channels as well as mass media, thereby paradoxically infusing virality into architectural features of end-to-end loops and encryption.

Technological and social conditioning notwithstanding, law and order objectives of the state can overrule encryption not only through traceability requirement clauses in internet regulations but also through brazen forms of surveillance, as evidenced in cases in the Global South when repressive governments normalize the practice of seizing mobile phones from dissenters and inspecting or downloading the data. At the same time, encryption has allowed dissenters to assert and safeguard

safe distance from state surveillance—a crucial communication infrastructure that has aided a large number of activist groups, from queer activists in the Middle East and North Africa to Black Lives Matter protestors in the United States and journalists critical of the regime in Rwanda (Moon, this volume). Ambiguities around encryption are pronounced in actual practices, when physical phone searches, for instance, spark a panoply of strategies to avoid or subvert such searches, puncturing a sense of security that messaging apps might proffer (Schumann, this volume). Indeed, as the chapters highlight, in several contexts, encryption is not an actively acknowledged technical feature or a centerpiece of how the messaging service is experienced and appropriated.

While affording protection to political critics in some contexts and disappearing as a feature of the media at the experiential level or inverting the logics of closed communication in other contexts, also by utilizing the platform's "narrowcasting" feature to send out messages to up to 1,024 individual accounts (WhatsApp 2023), encryption has nonetheless raised greater barriers for antihate and fact-checking initiatives, as access, storage, retrieval, and response to problematic content become more challenging and resource intensive. In authoritarian East African countries, security affordances such as end-to-end encryption serve as "window dressing that reinforce the perception of the surveillance state and make information verification more, and not, less complex" (Moon, this volume). Encryption thus entails new hurdles for scrutiny and verification, breeding innovative practices of "informal" fact-checking and gray interventions (Mare and Munoriyarwa, this volume). Even more, social intimacy of lived encryptions not only eases circulation but impedes correction, since users avoid calling out misinformation to limit social frictions, as evidenced by young users in South Africa not venturing into correcting the messages of elders in the group out of respect (Wasserman and Madrid-Morales, this volume). This stands in contrast to quasi-public forums where correction faces no such hurdles, although the effects of fact-checking and corrections are not guaranteed and can even backfire.

Finally, encryption in the diasporic contexts in the Global North offers pathways to craft alternative channels of communication distinct from majority dominated "mainstream" communication, affording

marginalized communities a way to articulate political matters and remain connected with families and publics in their homeland (Trauthig, this volume). Encryption here evinces, if only partially, the ideal of subversive speech within relatively well-guarded spaces.

We consider such multifarious unfoldings as "lived encryptions," holding vital significance, as the contributions in this volume illustrate, for how extreme speech spreads and entrenches in public discourse. As opposed to "encryption" as a descriptor of a distinct and bounded technological feature, "lived encryptions" foreground contradictions and multiple lived practices that surround the messaging application. By pluralizing the term, we emphasize that a contextualized understanding—of how closed architecture is broken open with political messaging and encryption disappears as a feature at the experiential level as well as the social sanctification of political content, the risk-reducing privacy feature, the collaborative potential of WhatsApp groups and so on—is critical to draw out the normative stakes of WhatsApp as a communicational condition that inspires intimacy at the starkly oscillating boundaries of bounded communication and networked action.

While "lived encryptions" highlight the multifarious dimensions and ambiguities surrounding encryption, it is important to note that the consequences of socially sanctified disinformation and rumors, riding on in-group socialites and cross-group message virality of WhatsApp, have been stark in the numerous cases we highlight here. For instance, mob lynching of Muslim minorities and oppressed caste groups has closely followed rumors, fake images, and misinformation circulating on WhatsApp in India while exclusionary narratives within xenophobic WhatsApp groups in South Africa have amplified hostilities toward people migrating from neighboring African countries seeking jobs and livelihoods (Sibiya and Gagliardone, this volume). Such impacts have contributed to what UNESCO has observed as a growing issue of online hate speech impacting the physical world, as evidenced by incidents of violence in Indonesia, Kenya, Bosnia, and other countries, where "cases of harmful yet lawful ('gray area') speech have often led to real-world violence" (Brant and UNESCO 2023, 45). The grave consequences of lived encryptions of WhatsApp in the form of physical attacks, state surveillance, and ideological fanaticism, while also offering spaces for

everyday conversations for a vast variety of social activities, stress the need for a global conversation and multiple angles of inquiries around this vastly popular messaging service.

Structure of the Book

We have organized the contributions under five sections tracing "lived encryptions" across different national and transnational contexts and with distinct focal points, which we outline next.

Politics of Divisive Messaging

This section will explore contextual social and cultural conditions that amplify the cocreation, consumption, and spread of disinformation and extreme speech on WhatsApp. The chapters will examine divergent practices surrounding WhatsApp, and how they fold into extreme speech as habitual, deliberate, and lived forms of discourse and meaning. The key focus will be on divisive, xenophobic, and partisan politics that draw on WhatsApp cultures, and how political campaigns deploy this messaging service to trigger animosities, panic, and violence. This analysis will also highlight examples of problematic content and networked dissemination patterns on WhatsApp.

In the first chapter, Parreiras draws on her ethnographic study of WhatsApp use in the favelas of São Paulo; she delves into a thick social world of moral values and shared anxieties and how the right-wing "Bolsonarista" groups amply instrumentalized them with their conservative discourse around "God and family." The far-right discourses on WhatsApp are embedded within peripheralized areas of the city "marked by different forms of material precariousness, such as insufficient sanitation, unemployment or underemployment, precarious housing, food insecurity, and difficulty in accessing health care." In such a context, also marked by heavy presence of the military and the police, WhatsApp use and political propaganda propel "local chains for spreading fake news, misinformation . . . and moral panics."

The next chapter turns to the anti-immigrant group "Operation Dudula" in South Africa. As Sibiya and Gagliardone show, WhatsApp serves as an important channel for the "informational and operational

objectives" of this group, mobilizing narratives against immigrants but also connecting with "broader civic activities." Users on such WhatsApp groups are more likely to distribute information incidents of crime involving immigrants from neighboring African countries with an objective to provide supposed fact-based information that could show the immigrants in a poor light. Xenophobic sentiment that shapes and binds such groups defies easy characterization of blind and misinformed ideologues. The authors reveal complex practices among majority Black South Africans who drive such xenophobic sentiments, showing how they embed their content within "fact-based" information as well as ironically express nostalgia for the apartheid and suspicion about pan-Africanism.

Highlighting right-wing nationalist messaging on WhatsApp groups in India, Udupa argues that WhatsApp's unique role in disinformation and vitriolic ecosystems in the Global South contexts of hegemonic politics lies not as much in the architectural features of encryption but around particular clusters of social relations it enters, entrenches, and reshapes. Describing this as "deep extreme speech," she suggests that it is "characterized by community-based distribution networks and a distinct context mix, which both build on the charisma of local celebrities, social trust, and everyday habits of exchange." This type of extreme speech, she argues, "belongs less in the problem space of truth or the moral space of hatred and unfolds rather at the confluence of affect and social obligation, variously inflected by invested campaigns."

In the final chapter in the section, Santos, Ortiz Fuentes, and dos Santos illuminate a highly tense political moment in Latin America: the writing of a new constitution in Chile and the final referendum that ended up rejecting the proposal. Based on quantitative and qualitative content analysis of discourses in a set of over three hundred WhatsApp political chat groups, their study reveals widespread circulation of conspiracy theories about election fraud. Although widespread, the circulation, they reveal, was "asymmetric." Right-wing groups were exposed to a more and larger diversity of misleading messages compared to left-leaning groups, and right-leaning WhatsApp groups were also "closely knitted" in that users were active in more than one group, thereby bridging the narratives across several groups. However, over time, hoax messages split over into left-leaning groups, suggesting how WhatsApp

circulation can lead to society-wide disruptions in political debate and citizen participation.

(Un)safe Spaces

In this section, studies explore subversive speech practices and political ambiguities on WhatsApp amidst perceptions of safety, disengagement, and fears of surveillance that at once surround WhatsApp. Across all the chapters, the authors highlight the tension between perceived safety of encryption and its affordances to forge networks beyond majority-dominated communication on the one hand and the risks, on the other hand, of exposure, mistrust, and ambiguities that actors negotiate in contexts of political volatilities. Exploring different communities of users—from political dissenters in a conflict zone and young users in the Global South to diaspora members in Western democracies—the studies show the deep ambivalence of WhatsApp as sites of community networking and (dis)informational sources.

Turning attention to invasive social media policing of the Cameroonian state, Schumann informs that "government suspicion against Anglophones existed long before" state practices of policing their WhatsApp and social media channels became widespread. Set in this context of a long-drawn conflict, "arbitrary phone searches" are common, and so are different ways in which Anglophones attempt to subvert everyday surveillance by leaving behind their smartphones at home or deinstalling social media apps—tactics that more often raise suspicion among government authorities since they begin to question why an affluent-enough person would not carry a smartphone. As Anglophone WhatsApp users vacillate between a sense of security around encrypted messaging and vulnerability to state surveillance, they also find themselves in awkward and potentially dangerous situations when violent images of killing surface unexpectedly on their phones, even as they hold on to the messaging platform as "private spaces of exchange" to "discuss negative experiences with state forces."

Wasserman and Madrid-Morales show that contemporary forms of disinformation campaigns in South Africa are rooted in "older histories of colonialism and postcolonial authoritarianism" and longer tensions surrounding ethnic and social polarization. In this context, young

WhatsApp users find themselves torn between actively attempting to counter misinformation they receive on their phones through various corrective actions and purposeful detachment from messages of this nature. What is safe and unsafe is shaped by thick social norms that surround young users' actions, especially in relation to how challenging false information in close-knit WhatsApp groups could cause problems in family relations.

Belinskaya and Rodriguez-Amat explore the understudied aspects of encrypted messaging about fleeing Russian migrants in Europe, revealing how bans on prominent "public" social media platforms in Russia have driven many users toward WhatsApp and Telegram. Offering thematic analysis of sampled WhatsApp groups used by Russian immigrant communities in Austria and drawing a comparison with Telegram, they show that misinformation is common within these groups. They argue that communicative processes of rationalization and legitimization significantly shape the exchange of such misinformation.

In her contribution, Trauthig shifts the focus to diaspora communities in the United States, arguing that WhatsApp use among Cuban American, Indian American, and Mexican American communities has allowed for alternative avenues to discuss contested issues and create counternarratives "outside of . . . majority-dominated public discourse." In this analysis, she compares WhatsApp with the features of community-owned media that offered an "alternative environment for inclusion and representation," crediting the "inherent subversiveness of encrypted communication" for this potential. Disputing extant evaluations of WhatsApp as a hotbed of disinformation, this study considers its ability to create channels for diaspora members to engage in "political talk" in ways that merit protecting such messaging channels as "safe news spaces" that could foster the ideal of inclusive democracy.

Infrastructure

Approaching WhatsApp as a sociotechnical architecture, the chapters in this section explore its shaping in relation to journalistic reporting and fact-checking and as a site for regulatory intervention and corporate moderation. This section also highlights how technical features of the messaging app are often overshadowed by vast human networks

deployed for political messaging in the Global South contexts, prompting the consideration of human networks as infrastructure in and of themselves.

The thrust of the section is on considering WhatsApp as physical networks that can enable the movement of information, ideas, and emotions in routinized, persistent, and standardized ways, akin to other physical infrastructures, and for this very reason, they are subject to state scrutiny. Building on key anthropological and communications scholarship on infrastructure, Moon examines the ways journalists in the authoritarian East African country of Rwanda use and are shaped by WhatsApp as an element of infrastructure: "a 'boring thing' that distributes justice and power in the background of visible work." WhatsApp use among journalists in Rwanda, she argues, is strongly impacted by far-reaching surveillance of the authoritarian state that operates by covert strategies of control over infrastructure, including messaging platforms.

Olaniran (this volume) expands the conceptual scope of "infrastructure" into what Larkin calls "people things" by showing how politicians in Nigeria "assemble a human infrastructure to create partisan environments and inflammatory messages to bolster their candidacy" (see also Nemer 2021). This very infrastructure of WhatsApp, however, has also been utilized by diverse groups of "bottom-up" fact-checkers in South Africa. Highlighting the "unofficial, uncoordinated, and unorganized process of verifying the factual accuracy of questionable content," Mare and Munoriyarwa show that informal fact checkers with no specific training in journalism or fact-checking rely on their "intuition, epistemic capital, social networks, media literacy skills, and investigative skills" to verify content circulated in closed WhatsApp groups. They use the conceptual device of "social correction" to explicate how WhatsApp infrastructure of closed chats partly necessitates "informal and provisional" fact-checking since it can venture into "gray spaces" with flexible correction tactics. Such tactics, they contend, can navigate the "unpredictability and uncertainty associated with circulation of mis/disinformation on WhatsApp groups" in contrast to organized fact-checking groups that encounter the stonewalling effects of encryption.

The difficulties of fact-checking WhatsApp are a part of the broader problem concerning the regulation and moderation of such instant messaging services. Content moderation on WhatsApp is hard to implement

at the message level, and platform moderation therefore becomes limited to placing restrictions on group membership and size, content labeling rather than removal, and mechanisms to decelerate message spread by limiting the number of forwards. Taking a closer look at such regulatory and moderation measures, Sinha considers WhatsApp infrastructure as a site of state and corporate intervention, discussing how regulation of WhatsApp has developed in India.

Method

The chapters in the preceding sections reveal the diversity of methods employed in current scholarship on WhatsApp. Each chapter uses one or more research methods, and across the volume, methodological approaches are diverse and multidisciplinary, ranging from quantitative computational methods to ethnography to surveys. Taking up the methodological question as a central concern, this section turns the focus on a major hurdle in WhatsApp research, which relates to the methods to access and store "private chat" data. Most studies have highlighted the methodological difficulties of studying encrypted messaging services because of lack of public API access, closed source code, sequential chats that do not allow key word searches, and the "private" nature of groups that requires group moderators to approve researchers' request to join (Barbosa and Milan 2019). Although Telegram and WhatsApp can be accessed via web browsers, and limited metadata can therefore be obtained with scraping, such methods raise ethical challenges and issues of data privacy, in addition to challenges of seeking approvals by parent companies and navigating company stonewalling.

Lack of data access has resulted in empirical blind spots in terms of assessing the volume of users exposed to extreme speech and the proportion of problematic content vis-à-vis the total corpus of exchanges. In view of severe limitations to data access, methods have emerged especially within computer science to develop web interfaces to encourage users to donate data and to organize such content in a "privacy-preserving manner on a large scale," but the adoption of such promising models remains nascent because of "serious privacy, legal, ethical, and practical challenges" including meeting the standards of data minimization and anonymization principles (Melo et al. 2019). At the same time,

epistemological implications of calibrating data donations with research goals are yet to be investigated.

Foregrounding vast ethical, practical, and legal challenges around accessing and analyzing conversations on WhatsApp and similar messaging applications, this section provides methodological pathways toward addressing them.

Inquiring into ethical dilemmas involved in "lurking" within private WhatsApp groups, Saka highlights the ethical risks of navigating access to such groups and specific challenges of carrying out ethnographic work. When researchers announce their presence on closed groups and seek to gain informed consent of participants, as he points out, they are likely to inadvertently change the "nature of conversations." Reflecting on interdisciplinary methods he adopted in his study on WhatsApp groups and anti-EU disinformation in Turkey, he suggests that methodological strategies for content and group analysis of WhatsApp should pay close attention to conditions set by national media environments and the platform's shifting policies.

Highlighting how WhatsApp data collection on a "large scale presents serious ethical and practical challenges," Chauchard and Garimalla take up the challenge of evolving methods that can gather, store, and analyze data in a privacy-preserving manner. On the one hand, data extraction is "technically easy" once a consenting participant extracts it for the research term. On the other hand, data gained through such means and subsequent analysis raise the risk of falling into the gray zones of the platform's community standards and prevalent legal protocols. The only way to address the hurdles, they contend, is by adopting a data donation approach, elaborating on different technical and practical steps involved in operationalizing such a methodological protocol.

Honing further into computational methods of gaining data access, Micallef, Ahamad, Memon, and Patil outline the challenges of "automating large-scale data collection from public WhatsApp groups." Identifying such challenges in the three activities of "discovering, joining, and maintaining membership" in public WhatsApp groups, they point out how they faced hurdles at every step of the process, for instance, when their code allowed their "phone number" to join but their inactive presence in the group raised suspicion and ended up being removed by the moderator. They emphasize that automating data collection for public

WhatsApp groups remains highly vulnerable to the evolving policies of the platform, which can shift quite suddenly and without notice for the researcher community. Cognizant of such challenges, they provide a detailed description of data gathering including setting up of the devices, extracting the WhatsApp SQLite database from the devices, manual identification of WhatsApp groups for research, semi-automated ways of joining public WhatsApp groups, and maintaining group membership.

In the final contribution, Bosch provides an overview of common qualitative methodological approaches in available scholarship on WhatsApp and political discourses, highlighting the significance of mapping this from the vantage point of the Global South. Noting that "the majority of research on WhatsApp and political activism originates from the Global South," her review finds virtual ethnography, interviews, and surveys of users to be the most common qualitative approaches. Approaching the internet as "place and text," the chapter reflects on the possibilities, challenges, and limitations of different methods that emphasize each or both aspects in relation to WhatsApp discourse. The chapter's central thrust is a decolonial approach to WhatsApp research, calling for historical awareness, self-reflexivity, and an ethics of care.

As the methodological contributions in the volume indicate, one of the key methodological tensions around WhatsApp research is what gets deleted from the dataset to preserve privacy, as the authors from different disciplinary backgrounds illustrate in this section and different chapters in the book point out with varying degrees of emphasis. While removing one-to-one threads for the sake of protecting privacy appears reasonable for large-scale analysis of WhatsApp content using computational methods, it raises the question of the profound ethnographic value of such conversations, as Schumann's study of Anglophone users in Cameroon and Udupa's study of deep extreme speech illustrate. Similarly, removing seemingly banal, insipid, and repetitive one-liners appears to be reasonable for computational methods for the "noise" they bring to the data, but ethnographers would consider them as valuable for examining the interactional dynamics and rhetorical devices that draw extreme speech into the everyday and the ordinary. While such methodological differences ultimately rest on the epistemological grounding of different projects, this volume highlights ways of bringing them in

close conversation, to highlight them not merely as problems of data selection and data cleaning but a methodological way forward for interdisciplinary collaboration. WhatsApp's intimate contexts of conversation with multiple contradictions underscore the value of ethnography in no uncertain terms, but precisely because of the political consequences of cross-platform velocity and virality they have induced in various contexts, computational and survey methods are vital to track their multiple trajectories. Taken together, the chapters in this section therefore underscore the importance of interdisciplinary research that combine different methodological approaches and epistemological orientations to highlight the different dimensions of WhatsApp in the communication landscape.

Policy

The final section aims to address the looming question around what to do with the complexities of encrypted messaging and extreme speech. The key challenge is to design regulatory and policy frameworks that can account for the diverse use of WhatsApp globally, and the dilemma of encryption in relation to protections it can provide to minoritized communities facing threat under surveillance and authoritarian regimes on the one hand, and the growing evidence, on the other hand, that affective and instrumental engagements around encryption have seeded possibilities for hateful exchange and disinformation. Regulatory models that have emerged include outright bans on encryption, limitations on the permitted strength of encryption, weakening of encrypted technologies, requirements for traceability of users, and requirements for back doors to be built into services and products to enable government access to information, and mandates for proactive monitoring of encrypted content. These proposals have emerged across jurisdictions including Europe, UK, Turkey, India, and the United States. The Indian government, for instance, introduced the traceability clause in the new Internet Intermediary Rules (2021), mandating the platforms to divulge information about the source and identity of viral messages in response to law enforcement requests—a measure that WhatsApp challenged in court (Sinha, this volume). Hence, while legally mandating access to encrypted messages appears to be an easy solution, it comes

with serious challenges including security and surveillance risks that are introduced by the existence of back doors, ways to balance the potential infringement on privacy, lack of clarity on processes to carry out regulatory mandates, and jurisdictional challenges that arise once a back door is introduced. In a long-drawn contestation with service providers, the UK's Online Safety Bill, for instance, shelved the proposal to inspect encrypted messages, admitting that the "technology to securely scan encrypted messages . . . does not exist" (Guest 2023). In addition, regulatory measures such as limiting message forwarding and even fact-checks have yielded mixed outcomes (Melo et al. 2020).

In this section, two fact checkers and one policy expert reflect on the challenges that WhatsApp has posed to civil society and regulatory efforts in curbing exclusionary extreme speech and disinformation. Through their daily navigations of WhatsApp discourse for fact-checking, Cayley Clifford from Africa Check in South Africa and Jency Jacob from BOOM Fact Check in India discuss how their meticulous and timely fact-checks face the danger of being drowned in the virality of polarizing and sensational content that spreads through community channels at a rapid pace. In India, for instance, WhatsApp messages morphed public awareness videos from Pakistan and bodies of little children shot in Syria to raise panic about alleged child kidnapping gangs and organ harvesting rackets. Published fact-checks could not mitigate the rapid spread of these messages, although fact checkers simultaneously alerted the platform to take action. Jacob points out that efforts to develop open channels for community collaboration are exhausting and resource intensive, as often they are flooded with spam messages, including cryptocurrency messages, while citizens who alert the organization are on edge to see immediate action. Such practical challenges are situated within a broader political climate of antiminority ideological politics in India, placing enormous pressure on independent fact checkers as they jostle between platform complicity and repressive politics.

Clifford similarly outlines the challenges as well as opportunities presented by WhatsApp for fact checkers, stating that it is not only difficult to access information within encrypted channels but also to counter disinformation since it is very likely that messages that flow through such channels are trusted. Highlighting their initiative *What's Crap on*

WhatsApp?, a podcast circulated as WhatsApp voice notes to debunk misinformation, she details how they established a tip line for users to directly alert Africa Check about suspicious content and prepare periodic podcasts to raise awareness. The organization vouches for constant interaction with subscribers for effective fact-checks that can also address backfire effects when fact-checks reinforce beliefs. Both Africa Check and BOOM run helplines for users to directly alert them on viral messages that appear dubious as well as collaborating with alert citizens who would contribute as "fact ambassadors" or "truth warriors."

Across political contexts, some of the key challenges that fact checkers face in relation to WhatsApp discourse have emerged from lax platform action and challenges posed by the platform's architectural design. In Brazil and India, for instance, Reis et al. (2020) discovered that even after popular fact-checking agencies had verified the information, misinformation continued to appear within public WhatsApp groups, as the platform lacked the capability to label previously fact-checked content. Consequently, fact-checking organizations, as Jacob informs, have demanded "in-platform mechanisms to fact-check high-volume forwarded messages without compromising encryption."

Outlining some of the major policy challenges concerning "coordinated harm" on WhatsApp, Scott Timcke stresses that platform governance alone cannot solve the problem since governments tend to divert attention away from social conditions and "their history of governing those conditions," thereby "co-opting platforms into projects which circulate narratives of hate." Furthermore, platform governance as a distinct policy measure is constrained in the case of WhatsApp because of the lack of public metrics to assess the flow and impacts of hateful narratives as well as the very structural conditions of digital capitalism that policies, including those of UNESCO, fail to address in terms of devising ways to decommodify global social media platforms. Pertinent to the discussion is what he defines as "a climate of stochastic violence," which refers to hazy boundaries between intentional actors, passive recipients, and unmindful onlookers that WhatsApp groups afford, raising the difficulties of striking a "balance around intent, primary audience, and harms." Considering these challenges, he suggests that policies turn to understanding broader sociological and sociotechnical processes to pin down the processes and consequences of harm.

The emphasis on ground realities that fact checkers and policy experts have articulated in this section returns to our opening argument about the need for multiorder analysis with field-based ethnographic approaches developing in close conversation with quantitative methods, and for contextualized understanding that considers users not as "targets" of analysis and policy but as historically situated actors who draw and imbue meanings within contradictory climates of encryption—of which some distinct forays have emerged in this volume. The global influence of WhatsApp as a communication platform, its increasing sphere of influence in the Global South, and its contradictory characteristics outlined in this volume call for further interdisciplinary and multimethods research, for which this volume has laid the foundations.