VIOLENT CONFLICT, TECH COMPANIES, AND SOCIAL MEDIA IN SOUTHEAST ASIA

Key Dynamics and Responses

The Asia Foundation

VIOLENT CONFLICT, TECH COMPANIES, AND SOCIAL MEDIA IN SOUTHEAST ASIA

Key Dynamics and Responses

The Asia Foundation

Violent Conflict, Tech Companies, and Social Media in Southeast Asia: Key Dynamics and Responses

The Asia Foundation
465 California Street, 9th Floor
San Francisco, CA U.S.A. 94104
www.asiafoundation.org

Contents

ACRONYMS

ARSA	Arakan Rohingya Salvation Army
ASEAN	Association of Southeast Asian Nations
HRW	Human Rights Watch
IPAC	Institute for Policy Analysis of Conflict
ICG	International Crisis Group
ISIS	Islamic State in Iraq and Syria
ITE	Information and Electronic Transfers Law*
MIDO	Myanmar ICT for Development Organization
M.MAS	Myanmar Media and Society
RUSI	Royal United Services Institute
UNESCO	United Nations Educational, Scientific, and Cultural Organization
UNHRC	United Nations Human Rights Council

* Translated from the original Indonesian acronym.

Preface

Compiling an overview of a complex and ever-changing field is not a simple task. New trends and technologies arrive fast, generating the risk that any study based on recent events is outdated even before it is released.

The field of new technologies and conflict is vast and wide-ranging, and the theme of this report — *Violent Conflict, Tech Companies, and Social Media in Southeast Asia* – emerged during the research process. Its aims are to inform future programming, to serve as a guide for policymakers and development actors, and to stimulate further practical research. We have sought to bring information to attention as rapidly as possible, while maintaining a rigorous, evidence-based approach to issues on which almost everyone with a smartphone has an opinion.

In ensuring that this report was completed before it became obsolete, some important concerns were necessarily deferred for future analyses. In particular, events outside Southeast Asia deserve closer attention, and the relationships between social media, violence, and gender discrimination are only addressed briefly here. Equally importantly, this report only begins to explore the scope for taking action to break the links between social media use and patterns of violence. The coming years are likely to see concerted efforts to promote oversight, monitoring, regulation, technical innovation, and other measures to moderate the negative impacts of social media on society and on political processes.

This report was produced by The Asia Foundation's regional teams for Conflict and Fragility, and for Technology Programs, with support from country offices across Southeast Asia. A consultant, Matthew Abud, developed the initial draft and conducted research for case studies. While every effort has been made to maintain accuracy, all errors, omissions, and positions are the responsibility of the authors.

Smartphone use is now widespread across Myanmar and all of Southeast Asia.

Overview

Online spaces are the new frontier of conflict and violence in Southeast Asia. With the rapid spread of mobile broadband, inexpensive smartphones, online social networks, and messaging applications, the internet has become an important space for civic dialogue as well as more harmful behavior. Violent groups and movements use digital tools to recruit new members, spread propaganda, and keep their organizations running. Conflict actors are increasingly adept at using the new capabilities and emergent properties of online spaces to communicate with target audiences, build coalitions, and counter opposing narratives with greater sophistication, precision, and guile. Such emerging applications of new technologies—spanning extremist insurgencies, majoritarian movements, and social protest and repression—have not yet been well documented or addressed in mainstream peacebuilding efforts across Southeast Asia.

Although Southeast Asian conflict protagonists do not yet use online platforms to the extent of some organized groups in other parts of the world, stakeholders should brace for change.

There are already many examples of violent actors in Southeast Asia harnessing the power of social media and messaging applications to shape violent conflicts. In the Philippines, violent extremists corresponded directly with ISIS leaders in Syria and used similar online tools and techniques as they planned, executed, and publicized the occupation of Marawi City in 2017. The governments of several countries have drawn on disinformation techniques originally pioneered in Russia to implement sophisticated information operations that spread messages via fake Facebook accounts and false news stories, in cases intentionally contributing to communal violence against religious minorities. Across the region, governments, activists, and nonstate armed groups actively use social media to gain domestic and international support for their cause.

Drawing on evidence from conflicts in the Philippines and Myanmar, and supported by other examples, this report provides a broad analysis of how violent conflicts in Southeast Asia are shaped or affected by online platforms and social media. Research was conducted in late 2019 using interviews, Asia Foundation case materials, online sources, existing analysis, and other available data. The report is one important step in ongoing efforts to stimulate further research, programming, and other action across the region, both within The Asia Foundation and among partners. It also offers a basis from which to address similar issues elsewhere in Asia.

This report does not seek to summarize all links between violence and social media. It focuses on the use of social media in order to shape violent conflict for political or ideological ends, covering three main types of conflict in the region. First, *subnational conflicts,* or armed conflict over control of a subnational territory within a sovereign state, are found in many Southeast Asian countries, usually in remote or border regions. Second, *violent extremism*, defined as the use of violence such as terrorism by religiously motivated groups, remains a significant concern. Finally, *violent mass movements and social protest* typically involve demonstrations or other collective action along with violent state responses, although some mass movements target minorities rather than the state. These three conflict types all involve large numbers of participants and are motivated at least partly by political aims. In practice,

many individual conflicts combine elements of more than one generic type of violence, are shaped by prevailing gender norms, and amplify existing inequalities.

Key Patterns

(1) Southeast Asia's conflicts are being adapted to online spaces.

Alongside rapid economic growth, cosmopolitan urban hubs, and tech-savvy young entrepreneurs, Southeast Asia in 2020 is also home to various forms of violent conflict.[1] As improving telecommunications infrastructure and inexpensive smartphones extend internet connectivity far beyond affluent elites, social media has become a venue for reconfiguring violent action in the region.

By adapting to social media, violent groups increase their ability to control the narrative of a conflict, manage operations, and recruit new members. They are able to explore new ways to raise funds, to gain mass support, to vilify opponents, and to organize acts of violence. Online spaces create new opportunities for both women and men who might not otherwise be active participants in violence to "join" and support conflicts with just the click of a button.

(2) Each online space offers unique and dynamic terrain.

Understanding how Southeast Asia's conflicts are adapting to the internet requires an analysis of how online spaces differ from zones of physical conflict. Conflicts have always been shaped by real-world conditions—physical terrain, economic incentives, political processes—and now they are also shaped by online geographies. In any given conflict zone, the form of online space is affected by local conditions: levels of literacy and the existence of common languages, the penetration and speed of internet connectivity, the prevalence of smartphones, local preferences for particular applications, and the nature of the conflict itself.

The technical features of a given platform or messaging application significantly affect its adoption and use by violent actors. This is true even for features that are offered benevolently and used peacefully in other contexts. For example, a violent, majoritarian, mass movement—the repeated targeting of Muslim communities by organized networks in Myanmar—was aided by algorithms, like those of Facebook, that create dense networks of ethnically homogeneous groups and that can be applied to prepare the ground for further polarization. The messaging application Telegram has been favored by some groups for the privacy of its end-to-end encryption. By allowing anything from one-on-one encrypted chats that keep no records, to group communications in the tens of thousands, Telegram met the needs of armed religious extremists such as the insurgent leaders who affiliated with ISIS in Mindanao, the Philippines. As new online platforms and tools emerge with distinct features and capabilities, motivated actors who are already used to switching between multiple platforms will quickly adapt.

With a bit of experience, savvy actors can exploit the contours of an online space to gain advantage over their opponents. Proponents of violence in Southeast Asia use a range of tactics carefully tailored to the online context: generating content intended to misinform, publishing coded or even explicit hate speech, attempting to silence opponents' online voices through harassment and threats, and posting inflammatory messages from high-profile leaders that are then shared widely by followers. Even more sophisticated methods are also used, including "sock puppet" accounts with false identities, "bots" that create the illusion of a mass movement, and websites that mimic authentic news outlets but publish false information. Some of these techniques were first developed for influence operations, backed by states outside Southeast Asia and by cyberwarfare combatants, before being picked up by local groups.[2]

(3) Conventional peacebuilding, mitigation, and research efforts have not kept up with online interactions and their implications for conflict.

An examination of existing peacebuilding and conflict mitigation programs in Southeast Asia reveals few civil society initiatives that focus on the online dimensions of local conflicts. A large proportion of these programs have been implemented in Myanmar, a country which opened up to civil society initiatives just as the internet was taking off, and where the links between violence and online space have received intense international attention, including widespread media coverage of Facebook users' roles in setting the stage for majoritarian movements. Nay Phone Latt's *Panzagar* (Flower Speech) campaign—a counterspeech effort encouraging positive, peaceful social-media use—is a prominent early example of attempts to mitigate online polarization and incitement to violence.

Gauging the effectiveness of these programs is complicated, and reliable evidence is scarce. Investigations by The Asia Foundation in the Philippines and by the Institute for Policy Analysis of Conflict (IPAC) in Indonesia are two of the few major research efforts focused on the online drivers of violence and conflict in Southeast Asia.[3] Beyond these studies, there is little data or research on this rapidly changing environment.

Government-led conflict mitigation often involves the policing or militarization of online spaces, a fraught scenario when the military and police may already be perpetrators of violence. In Southeast Asia, where conflicts frequently stem from grievances against an overbearing or authoritarian state, policing the internet can exacerbate the cycle of violence and severely limit civil liberties. Public institutions have a duty to address online threats of violence and a right to publish their own media through online channels, but it is important that this work does not cross the line into censorship, propaganda, or the closing of space for civil society.

(4) Platform companies are unprepared to respond to the increased use of their products by violent actors in Southeast Asia's local conflicts.

Despite real progress in the past two years, companies that operate online platforms and messaging products—such as Facebook, Twitter, Telegram, Google, and others—have struggled to develop internal mechanisms to manage the growing problem of hate speech,

abuse, and the activities of organized violent groups. As public awareness of violent actors' online presence has grown, and as the notion that platforms have limited responsibility for the content that they carry becomes less acceptable, these companies have had to respond more quickly and proactively, to the point that they now employ tens of thousands of reviewers and policy analysts to decide what should be permitted on their platforms and what should not. Still, evidence from around the world suggests that their methods for moderation and control often fail, especially in areas where public pressure and oversight is not as strong.

Southeast Asia has many characteristics that make online content moderation difficult. The landscape of conflict includes many localized and small armed groups, some of which also act as community benefit organizations, ethnic or religious networks, and political parties. The tactics these groups use vary widely and change rapidly depending on local demographics and operational needs, precluding a one-size-fits-all response. The relatively little purchasing power of some countries in the region, along with their linguistic and cultural diversity, means that the attention of corporate policy teams is spread thinly. In addition, Southeast Asian state actors may themselves be protagonists in violent conflicts, and may attempt to use their lines of communication with platform companies to silence the voices of peaceful dissidents.

In recent years, several platform companies have expanded their conflict monitoring teams and begun more regular outreach to experts.[4] This is a positive step, but local civil society organizations working on the ground in conflict-affected areas still point out numerous inconsistencies, grey areas, and blind spots in both policy and enforcement, as well as delays in gaining the attention of platform moderators in urgent situations.[5] Local peacebuilders, mediators, and conflict mitigation experts could be important partners to platform companies in drafting policy, raising awareness, and enforcement, but such a relationship requires consistent engagement and trust. Corporate structures that concentrate policy expertise and oversight at the regional or international level also limit their effectiveness in Southeast Asia's local conflicts.

Covid-19, Conflict, and Social Media

Disasters and crises—floods, earthquakes, wars, and of course pandemics—shine a harsh light on society's problems and exacerbate existing inequalities. Covid-19 is no exception. It has fed prejudices, sparked violence, and shown where governments are not trusted or armed groups hold sway on the ground.

Infectious diseases do not respect political boundaries or the frontlines of war, and the need for joint action to counter the pandemic has led in some cases to new cooperation on the ground. Sadly, governments and armed groups have also looked to use the pandemic as a chance to gain territory or to win the propaganda war. In Afghanistan, the Taliban have been quick to showcase their health measures on social media and through traditional channels, while continuing assaults against government forces.[6] The online propaganda war between insurgents and the military has continued in Myanmar's Rakhine State alongside intense fighting on the ground. ISIS and other extremist groups have also sought to spread propaganda and promote recruitment on the back of the pandemic.[7]

Fear of the pandemic has unleashed waves of ethnic prejudice in many countries as people have turned against minority groups or migrants for allegedly spreading infection. In some cases, specific "super-spreader" events have triggered online hate speech and led to physical attacks. In Malaysia, social media channels have reported a spike in posts denigrating unregistered migrants, particularly Rohingya refugees.[8]

These identity-based tensions have not escalated out of control, however, even in countries with a history of unrest. Since the violence was triggered by fear and rumor, and in most cases was not actively stoked by political mobilizers and their online networks, it has mostly remained in check, and public order has been upheld or rapidly restored. What is more, as the pandemic has become increasingly widespread, its association with specific groups has weakened.

Governments have tried to take advantage of the pandemic to introduce new online surveillance measures that exceed what is needed to trace suspected carriers and that may reinforce existing inequities and norms. In Cambodia, for example, a new emergency law grants President Hun Sen extensive new powers to carry out surveillance of telecommunications and to control social media. Introduced alongside increasing suppression of opposition voices, the new law threatens long prison sentences for violations of loosely framed emergency measures.[9] These tensions and challenges will be further complicated by the economic impact of Covid-19, which may fuel disharmony in regional hotspots where resources are scarce and contested. The strong economic headwinds will continue to spread across the Asia-Pacific region for upwards of two to three years, stoking frustrations and divisions in locally contested areas and raising the likelihood of national-level protests in many countries.[10]

Recommendations

In light of these patterns, stakeholders seeking to diminish online contributions to violence in Southeast Asia should consider several ways forward:

Nongovernmental groups, academics, public and private funding organizations, and platform companies should strengthen independent monitoring and evidence-based research into the online dimensions of violence.

- *Independent data collection and monitoring:* CSOs, donor agencies, and academics can support more informed policymaking and rapid response to conflict by developing and tracking sex-disaggregated, data-driven indicators of the online contributors to violence, such as network polarization, online hate speech, and online recruitment and advertising by known violent groups. Donor agencies can increase the ability of independent local groups to collect, process, and publish this data.

- *Attention to local context:* Analysts and scholars can help build a strong evidence base on the use of online platforms in local conflicts by paying specific attention to how local, contextual factors influence the online environment. This should include further research on and support for the roles of women and women's networks and movements as active agents in both conflict and

peacebuilding discourse online. Research efforts should be based at trustworthy institutions, and civil society organizations and other local partners can support the design and implementation of fieldwork and analysis. Regional development actors and NGOs can support this research and promote international cooperation among CSOs, researchers, and online platforms. They can also ensure that relevant findings are shared openly and presented in policy forums.

Governments, with consultation and external support where needed, should develop balanced policy frameworks that promote inclusion and shared norms.

- *Policy reform for transparent, norms-based governance of online platforms:* Governments should align policies governing online speech and communication with frameworks that preserve privacy and freedom of expression. Governments can use policy levers to hold platform companies accountable to commitments made in the 2019 Christchurch Call, including transparency reporting, enforcement of community standards against violence and conflict, and support for peacebuilding.[11] Governments can develop and maintain consistent lines of communication with platform companies and, when new rules are proposed, consult widely to engage the community.

- *Creation of new norms and standards through consultative approaches:* Governments can work domestically, regionally, and globally to develop new norms and doctrines that condemn the use of social media platforms for the instigation or organization of violence by any actor. Open consultation and independent oversight will help to address concerns that governments will use new standards to limit freedom of expression, stifle critical voices, or suppress peaceful opposition movements.

Platform companies should refocus policy development and monitoring to support local leadership and public engagement.

- *A voice for local stakeholders in platform policy development:* Platform companies should establish consistent dialogue and auditing mechanisms that give a voice to local stakeholders, and hire local people and others with on-the-ground expertise to guide this process, including women and women's networks. The need to invest in and adapt to the diversity of languages across the region can go hand-in-hand with greater voice for local stakeholders.

- *Support for local capacity building and training:* Platform companies can support capacity building and training to enable local-level and multilingual reporting of hate speech, misinformation, and suspected disinformation operations, thus improving the quality of intraplatform reporting and building user trust.

All parties should support innovation in peacebuilding and conflict-prevention practices.

- *Digitizing conflict reduction:* All actors, including governments, international agencies, and NGOs, should aim to understand the specific role and influence of online activity as a core part of their conflict prevention and reduction efforts. Peacebuilders and mediators should apply their deep understanding of local (offline) contexts to develop appropriate peacebuilding methods for online spaces, with rigorous and ongoing evaluation to determine new best practices.

- *Tools to support analysis of the digital terrain:* Practical guidance, online resources, and training could help agencies navigate the digital field. Smaller and more traditional organizations can learn from the experiences of more tech-savvy peers and partners. Governments also need support to enable them to engage in peacebuilding and to ensure that security measures are not repressive or in contravention of human rights. Regional forums such as ASEAN or more specialized bodies can play a valuable role.

- *Empowering peacebuilding groups:* Governments should recognize and support the ability of civil society groups to address key drivers of conflict at all levels. Local groups can also actively engage in dialogue around relevant technology policy, especially encryption, cybersecurity, and policing of the internet. Governments should particularly commit to supporting the participation of women's groups, networks, and movements in peacebuilding and peacekeeping, per the commitments made in UNSCR 1325.

- *Alignment with local research activities and open data collection:* Open-access violence monitoring tools and databases improve our common understanding of conflicts around the world. Existing violence monitoring systems run by local and national research units or civil society organizations can be harnessed to grow the evidence base around social media and conflict. Local actors conducting online counterspeech campaigns, training and awareness-raising, or data collection can also be brought into broader global research efforts on social media and conflict.

Devastation in the center of Marawi, the Philippines. Social media was harnessed by armed groups involved in the urban siege of 2017.

Background

Online platforms

In simple terms, online platforms facilitate interactions between two or more actors online. Examples include various types of social media, marketplaces, and communication services. Rather than creating, acquiring, or delivering original content directly, these products connect people to each other, linking customers and merchants, writers and readers, or large networks of friends and acquaintances.

To succeed commercially, these platforms need to connect large numbers of people. Thanks to advances in cloud computing, the prevalence of cheap smartphones and mobile broadband, and common software standards and frameworks, online platforms can grow incredibly quickly with relatively little overhead. For example, Reddit, a social news and discussion platform, had 430 million monthly active users and 1.7 billion comments in 2019, but fewer than 500 employees work at the company.[12] In most cases, online platforms can scale up to serve very large audiences while relying on their users to generate original content, manage online communities, and create the product's value to others.

The design of an online platform has strong effects on how it is used and who its audience is, in ways that may be positive, negative, or neutral. Twitter's 280-character limit and Instagram's photo filters make it fun to post content and share what you see, but some studies indicate that these design elements can also decrease the civility of online discourse (in Twitter's case) and lead to negative body image for some users (on Instagram).[13] YouTube's searchable video database and personalized recommendation algorithms encourage deep exploration into a user's areas of interest, but when those interests overlap with extremist views, it may accelerate radicalization.[14] Understanding of these unintended consequences of online-platform design is still limited by a lack of public data and the dynamic nature of platforms' uses.

Many social networking platforms encourage users to share publicly with the world, and they facilitate global discussions of all kinds of political and social issues. Other platforms—messaging applications in particular—take the opposite approach, and instead support closed discussions among small groups. The critical aspect of these platforms is not necessarily reach, but privacy. This makes end-to-end encryption, which provides strong privacy guarantees, increasingly popular among privacy-conscious citizens as well as violent groups attempting to avoid detection by law enforcement.[15] Debates around online privacy and encryption have pitted digital rights activists and technology companies against police and state security agencies for decades now.

The rise of online information warfare

New technology gives violent groups new ways of operating. With just a cheap smartphone and an internet connection, groups can conduct fundraising activities, recruitment efforts, propaganda initiatives, and precisely targeted disinformation campaigns.

Studies focusing on conflict and the use of social media can be traced back to early research on the organization and communication strategies of nonviolent protests such as Occupy Wall Street and the Arab Spring.[16] Early theories regarding the use of social media and the state's response were then tested and explored during subsequent mass-mobilization events that at times tipped into sporadic violence or escalated to significant levels of violent conflict. Such was the case in Syria, where the civil war and the rapid rise of ISIS has, at least in part, been traced back to the growing use of social media, mobile phones, and other tools to create and share messages online, recruit members, share propaganda, and promote movements globally.[17]

ISIS used social media platforms strategically to achieve massive, instantaneous reach and visibility, allowing the group and its affiliates to recruit followers from far beyond the immediate conflict area.[18] Various online subcommunities emerged, engaging with the content in a variety of ways without direct action by the organization.[19] Mainstream media also consumed and republished content that ISIS circulated online, in some cases enhancing its propaganda impact.[20] Today, despite the declining significance of ISIS in contemporary conflicts, much of group's messaging and content remains online, as material is circulated over and over again across multiple formats and channels and used to sustain current and future jihadist movements.[21] In addition, the techniques used by ISIS leaders that led to their success online have been directly or indirectly reproduced in other conflicts.[22]

Both governments and nonstate armed groups in Southeast Asia continue to learn and refine their operations based on what appears to work in other contexts. Some jihadist combatants in Mindanao received explicit training by ISIS operatives in Syria, which informed their use of Facebook and Telegram during the 2017 occupation of Marawi.[23] Meanwhile, the Myanmar government's armed forces (known as the Tatmadaw) apparently learned techniques of online disinformation through longstanding links with Russia.[24] Elsewhere, too, governments have shown adaptability and responsiveness that are new to the region, drawing from an increasingly popular playbook of cyber-influence operations.

NARRATIVE GOALS AND COMMUNICATION TACTICS

Violent actors use online platforms to pursue the strategic goal of controlling a conflict's narrative. A narrative is produced through the aggregation of many pieces of content that confirm an overarching theme. Communities of violent actors select content to fit a particular narrative, to justify the use of violence, and ultimately to present an overarching worldview. This worldview aims to explain what is taking place to target audiences, to the point that the protagonist sets the terms for understanding the conflict.

Across Southeast Asia's conflict zones, violent actors claim to be in an existential struggle to defend their group or fulfill their group's destiny. Social media allows them to reinforce this narrative in powerful ways. Online platforms allow users to present an endless feed of targeted material. This sheer volume of content, drawn from both national and international sources and displayed to audiences that have already been self-selected into networks with shared worldviews, can appear to provide undeniable confirmation of a particular narrative.

Tactics and content

The tactics deployed to communicate these narratives, whether by a central leadership or more organically by a wider base, are key to their overall impact. Many of these tactics are well-worn propaganda techniques that gain extra potency from the peer-to-peer nature, immediacy, and volume of content on contemporary social media. Communications efforts can range along a spectrum of overlapping objectives:

- Emphasizing norms or exacerbating social divisions
- Demonizing or denigrating out-groups as enemies
- Advertising an actor's success, strength, or conviction to encourage supporters
- Smearing, suppressing, or simply drowning out alternative voices
- Legitimizing violence and calling for further violence
- Calling for supporters to carry out specific attacks or battles

Some actors—such as majoritarian chauvinist movements—can pursue these objectives in sequence, building in intensity over time. Tactical objectives can also change in response to events, jumping forward or looping back as circumstance dictates. Materials designed to achieve these objectives can run the gamut of modern social media content: text, images, memes, hashtags, video, and more. Different types of content can serve tactical objectives:

- Documentation, especially in video, of a violent actor's dramatic deeds
- Hate speech in any format

- Rumors of impending attacks, often propagated as disinformation, followed by calls for a preemptive response[25]
- Confirmed, rumored, or false reports of attacks or unjust acts, with calls for local retribution

Online Narrative-Building by Extremist Leaders in Myanmar

Majoritarian chauvinism in Myanmar invokes a narrative that the Buddhist religion is under existential threat from Muslims across South and Southeast Asia. This draws on historical tensions from the colonial era and relies on old discriminatory tropes which have been reinvigorated by recent identity-based polarization and conflicts.[26]

Online platforms allow this narrative to be propagated widely and with great force by sharing an enormous range of material—including ISIS videos and content from other Buddhist-majority countries—and describing it as "irrefutable proof."[27] The narrative allows local events such as interpersonal tensions or business disagreements to be reframed as a local example of a greater threat, thus drawing in many more people who would not otherwise become involved, radically escalating the original incident. This process has occurred repeatedly across Myanmar. In July 2014, violent, multiday riots spread across the city of Mandalay, stemming from the accusation that a Muslim tea shop owner had raped his Buddhist employee. The accusation was later found to be a false claim motivated by a business rivalry and patriarchal gender norms.[28] Still, through amplification on Facebook by prominent extremists, the story was quickly reinterpreted as an essential struggle.[29] Two people were killed, more than twenty were injured, and a mosque was burned in the violence.

Social media platforms enabled extremist leaders from the Buddhist network known as Ma Ba Tha, the "Patriotic Association of Myanmar,"[30] to increase their prominence and influence—in particular the monk Wirathu, who has achieved national and international notoriety with his polarizing online speech and political demands for discriminatory policies.[31] Such content has spread through Facebook, YouTube, and individual websites, along with DVDs, adhesive stickers, and other hard-copy material. Wirathu gained hundreds of thousands of followers who further disseminated his posts. The arguments and rationale that he promoted were used to justify discrimination and violence.[32] A United Nations fact-finding mission on Myanmar cited hate speech on Facebook as a key factor provoking acts of violence, both in Rakhine State and elsewhere across the country.[33]

Consolidating narratives

Narratives can be launched, strengthened, and consolidated by dramatic events that draw powerful publicity and reach target audiences. Conflict actors may stage high-profile attacks in which online dissemination and mass media news coverage are more important than gains on the ground. Others may not launch attacks themselves, but may take advantage of attacks by others to turn a narrative in their favor. In both cases, features of social media again give extra power to narrative-setting strategies.

Consolidating Narratives During the Occupation and Siege of Marawi

On May 23, 2017, local jihadist groups began a battle against government forces in the city of Marawi in the Bangsamoro Autonomous Region in Muslim Mindanao, the Philippines. The siege was predominantly led by brothers Omar and Abdullah Maute's Maute Group and Isnilon Hapilon's Abu Sayyaf. Both Hapilon and the Mautes had been involved in separatist militias for years, most recently working within small groups that had broken away from the Moro Islamic Liberation Front (MILF) and the Moro National Liberation Front (MNLF) when they began peace talks with the government. The siege of Marawi lasted five months, causing over 1,000 deaths, the displacement of hundreds of thousands of civilians, and widespread destruction across the city from government airstrikes and artillery and the razing of churches and public buildings by militants.[34]

Over several decades, Mindanao's many conflicts have been primarily based on local and regional grievances, with violent actors mapping onto clan and language groups, but, increasingly, some local armed organizations have stressed elements of international jihadism alongside local grievances, as was apparent during the five-month siege of Marawi.[35] Presenting local conflicts as part of an international struggle had the potential to attract young fighters disillusioned with current leadership, as well as possibly uniting militant groups previously divided by clan and other factors.[36] In the years immediately preceding the siege, several militant groups pledged allegiance to ISIS leader Abu Bakr al-Baghdadi in Syria.

In line with the highly visual propaganda techniques developed by ISIS, the allegiance-swearing by the Maute Group and Abu Sayyaf leaders was formalized in video declarations disseminated online. Video sites like YouTube constantly monitor and remove such content, but extremists have consistently found ways around the filters—for instance, by sharing "private" videos only accessible by direct link or just being quick enough to copy a video onto a range of other, less-monitored backup sites before the original is taken down.[37] Throughout the Marawi siege, militants produced sophisticated, in-house video coverage, including drone footage of the ruined city and pleas from prominent hostages.[38] This arresting footage helped draw attention to the conflict and promote a particular narrative.

The Telegram platform kept Marawi militants in contact with large and committed audiences across the globe. Field reports were rapidly translated and shared in multiple languages, including English, Arabic, Indonesian, Turkish, and German, as well as Tagalog, strengthening the international constituency for the combatants. Coverage sought to highlight the militants' successes in battle, but also to criticize government airstrikes that damaged large parts of the city. Meanwhile, the Armed Forces of the Philippines conducted their own social media operations to discredit these claims and promote their humanitarian efforts in the city, battling the militants both online and offline.[39]

The extent of foreign material and financial support for the Marawi militants was somewhat limited and hard to trace, and it's likely that local contributions played a larger role than reported, but the narrative-building impact of the siege was significant.[40] Through chat groups on Telegram, Maute Group and Abu Sayyaf fighters reached out to global networks

of ISIS supporters and pushed hard for official statements via an ISIS-affiliated propaganda account, the Amaq News Agency, which itself operated primarily on Telegram.[41] In online chats collected by IPAC, Maute Group members could be seen correcting Amaq statements that they felt did not accurately reflect their narrative of the conflict, while ISIS propagandists pressed Filipino fighters to keep their communication "on brand"—for instance, by encouraging adherence to a pan-Islamic group identity rather than local allegiances:

> *We advice [sic.] our beloved brothers who sacrifice their time and support their State. Don't use "Maute" "Abu Sayyaf" or any group name that pledge allegiance to Amerul Mu'mineen Abi Bakr al-Baghdadi (hafidhahullah). Instead call them, Soldiers of the State (Junudul Khilafah), IS fighter, IS of East Asia.*[42]

Organic action

In cases of social protest that involve or are met with violence, established narratives can also be mobilized "organically" or "virally" through online, grassroots networks. Online calls to action can be especially effective when narratives build on the grievances of a tightly connected identity group. Widespread and diverse participation in a movement enables protests to spread without structured leadership or financial backing. Prominent recent examples include protests in Hong Kong, by Papuans in Indonesia, and in Myanmar's Rakhine State, where the rapid expansion of the Arakan Army has been fueled by extensive grassroots support.

Such protests may emerge suddenly and lack a formal organizational structure, and yet they often develop a consistent visual identity online.[43] They draw on recent histories of online activism—"clicktivism"—among social media users, often around low-risk subjects that do not contest dominant national or religious ideologies.[44]

Grassroots mobilization may draw an online response from its object of protest, especially from state agencies with greater resources. One common government measure is to cut off the internet to restrict mobilization and information sharing. By June 2020, the Myanmar government had restricted internet access in parts of Rakhine State for over a year in response to ongoing conflict.[45] Meanwhile, ethnic Rohingya who fled Rakhine State faced a separate communications blackout in refugee camps across the border in Bangladesh.[46]

In both Bangladesh and Myanmar, humanitarian agencies complained that the blackouts restricted access to health information, undermining efforts to limit the spread of the coronavirus. Human rights activists have interpreted the internet ban in Rakhine State as a form of collective punishment, hindering access to information and media that are regarded as an essential part of contemporary daily life and economic activity.[47]

Pro-state responses and propaganda

Beyond providing a theater for organic action, social media platforms are open to active manipulation. Those with greater skills and greater resources—financial resources, institutional powers, or numbers of supporters—have significant advantages, not only in promoting their own narrative, but also in drowning out alternative or local perspectives through sheer volume of content.

The online activities of state agencies as conflict protagonists deserve particular attention. The responsibility of every state to uphold and protect citizens' rights makes violations more deplorable—including violations using social media. What is more, states typically have far greater resources than nonstate actors to pursue their online goals. The scale and impact of state actions are often purposely obscured, but the case studies presented below paint a persuasive picture. They demonstrate how states were using newer technical measures to manipulate online discourse, at least until actions were halted by online platform companies' increasingly busy anti-abuse teams. State agencies have also employed more conventional legal measures, such as sedition laws in Thailand and Malaysia, to control critics online.[48] Laws that restrict online speech in this way can become tools of repression for state agencies.

The Myanmar Government and Social Media Operations

Social media and websites that spread false information as if it were news, sometimes mimicking the URLs of legitimate local news sources, are found in several countries in Southeast Asia. Credible evidence has linked many such sites to government activity.[49] The authors of one respected study commented:

> *The implications of this go beyond these specific case studies…to reflect the way global social media platforms and the power of international narratives are increasingly becoming a key battleground in local conflicts.*

The role of governments in defining and manipulating narratives is of particular concern. Myanmar is commonly cited as a country where such actions have been associated with subsequent acts of violence. The country's most significant and well-known influence operations have occurred on Facebook, although Twitter and other platforms have also been used.[50] Hundreds of pages operated by the military, including accounts openly attributed to top generals, were taken down by Facebook in the second half of 2018, following a UN fact-finding mission's report that the platform was being used to incite hatred and encourage violence.[51]

This major propaganda effort, which continued online for several years, involved managing pages under a variety of guises, including using the names of well-known celebrities. Material from those pages was further distributed through troll accounts.[52] By identifying and removing these accounts, Facebook confirmed active manipulation of the system. Many of the most popular pages that were taken down by Facebook had nonpolitical titles,

including "Let's Laugh Casually" and "Young Female Teachers." In designing this content but making it seem to come from citizens themselves, the pages drew on techniques practiced in Russia, and on Myanmar's own long experience with conventional propaganda.[53] Facebook's wide reach, audience engagement, and ability to be gamed by fake accounts allowed the promotion of negative and hostile attitudes towards Rohingya people in particular.

The content of these Facebook pages was similar to the content of Ma Ba Tha pages, but the tactics differed. Whereas Ma Ba Tha leaders shared their content openly, the government page used fake accounts and pseudonyms. This partly reflects the different positions of the two institutions: Ma Ba Tha extremist leaders benefited from their status as high-profile Buddhists, which was boosted by their online content, whereas the banned Facebook pages were meant to seem like grassroots content, making subterfuge imperative.

Organizing Operations

Social media not only shapes opinion; it can also mobilize resources. In other words, online platforms can influence physical action. In some cases, this expands capacities that were already there; in others, it enables actions that would otherwise be impossible.

Recruitment

Violent actors' organizational needs differ, and so does their use of social media to meet them. The recruitment of supporters to campaign or fight for a cause is a key step for insurgent or terrorist groups, and social media content often includes calls to join such causes. Individual recruiters also try to identify and reach out to sympathetic or susceptible individuals online.

Studies of extremist right-wing and jihadist movements in the West show a clear pathway of recruitment beginning online and passing through various stages before any physical contact or actions take place.[54] But this does not appear to be the pattern in Southeast Asia, where effective recruitment initiated purely online is negligible according to available research. Instead, the role of social media is almost always an extension of preexisting networks in the physical world.

More generally, online connections sustain jihadist communities in isolated locations. This is perhaps the broadest and most significant function of social media platforms: rather than playing purely instrumental roles such as facilitating militant recruitment and the planning of violent acts, they help sustain the jihadist social milieu.[55]

After many years in which the absence of "lone-wolf" and suicide attacks marked a key difference between Indonesia and many countries in the West, a small number have occurred, including suicide bombings by women. These were directly inspired by the battle for Marawi, among other factors.[56]

Online Recruitment Patterns in Mindanao and Indonesia

Armed groups in Mindanao have tried various ways of online recruiting: direct messaging to known individuals, spamming closed groups where a likely recruit is a member, and reaching out to potential recruits in public groups to invite further virtual or real-world contact.[57] On the same Telegram groups where jihadists worldwide followed and supported the battle for Marawi in Mindanao, ISIS Central in the Middle East urged would-be fighters, particularly in Indonesia, to go to the Philippines if they couldn't come to Syria. It is hard to tell if these specific calls actually motivated people to join the fight in Mindanao after the siege began, and as time went by, ISIS group administrators on Telegram grew concerned about possible infiltration and began to warn against revealing sensitive information. Even so, online engagement is a key aspect of local and international recruitment.

Indonesian jihadists are part of a more complex and evolving dynamic between online and offline networks. Initial face-to-face contact remains essential to most recruitment. Often this happens first through forums such as study groups. Social media then allows this engagement to continue.[58] Platforms play crucial roles not only in recruiting individual supporters, but also in fostering new constituencies.

Perhaps the most vivid demonstration is the alleged radicalization of some Indonesians working as maids and in other professions in Hong Kong. This process evolved over several years, initially through online and physical gatherings for religious teaching. As the gatherings grew, more radical teachers made their way to Hong Kong. Again, platforms allowed ongoing contact between teachers and students. Eventually, radicalized individuals sustained their own communities and connected to a wider network of online jihadists.[59] Several Indonesians from Hong Kong are thought to have since taken on multiple noncombat roles, becoming effective propagandists, donating and raising money for the cause, and organizing and funding would-be ISIS fighters from Indonesia to travel to Syria.

Enabling women's participation in conflict

Traditional studies of the role of women in conflict have focused on their secondary roles as supporters or influencers of male participation in violence, and on women as victims of violence. Social media platforms are increasingly used to enable trafficking, gender-based violence, and other forms of exploitation of women in conflict-affected areas.[60] But evidence is now emerging of women participating actively in violent movements. Social media has enabled women to become directly involved through pathways that didn't exist before.[61] Though not usually involved as combatants, women are known to have used their online presence and networks to contribute to tactical support, propaganda, and partisan tasks.[62] Quantitative analysis of some extremist networks indicates that, on average, women are better-networked within online communities than men and play important roles as connectors and brokers. Among pro-ISIS groups on the Russian social media platform VKontakte, the proportion of women in a group is directly correlated with its longevity.[63]

Research on this phenomenon in Southeast Asia remains limited to a few cases, particularly Indonesian religious extremists. Several Indonesian extremists were early adopters of online platforms, including websites and listserves, which became significant in building their movement.[64] However, these were centrally managed by a few individuals—that is, men who also acted as gatekeepers, deciding who could join activities and discussion, with women largely excluded (some Indonesian women participated online under male pseudonyms as late as 2009). Social media's peer-to-peer connections completely bypassed this control, enabling many women to become directly involved and connect to other, like-minded individuals.

Social media has increasingly enabled women to take on several roles in these circles: dedicated individual propagandists, fundraisers (through both seeking and making donations); recruiters of others to the movement (usually in combination with physical world connections); and facilitators of international travel for jihadist fighters. Another significant feature is online dating and marriage with male jihadists, even under circumstances where the couple never physically meet. Such marriages can add to the significant influence of some women within the movement.[65].

Social Media, Women, and the Growth of Ma Ba Tha

Ma Ba Tha is an extensive Buddhist community network in Myanmar that supports religious and community-based improvement. It has also been used by some of its leaders as a vehicle of strong anti-Muslim sentiment. Extremist leaders within the group achieved an extremely high profile by posting hate speech on Facebook, and the platform became a major way for audiences to engage with their agenda. Ma Ba Tha has grown in breadth and reach with extraordinary speed since its establishment in 2013, and the group has largely abandoned a formal membership process and instead operates through phone-tree networks.[66]

Women play important roles within Ma Ba Tha. Many women join the network because of its social welfare activities and may well disagree with its more divisive leaders. (In fact, many cite feminist aims as a reason for joining.[67]) Some women have also become fervent supporters and promoters of its majoritarian chauvinist agenda.[68] In either case, social media has dramatically lowered the barriers to entry for women. While women are traditionally active in social and religious networks across Myanmar, they have taken on new roles with Ma Ba Tha's rapid growth. The online sphere provides access to a public forum beyond the local level, bypasses traditional monastic leaders, and provides a direct channel for divisive messages or incitement to violence.

Fundraising

Another critical role supported by social media is fundraising, which often consists of seeking donations from supporters. In the Southeast Asian cases where research exists, financial transfers are frequently organized online, although the money appears to be sent using conventional commercial services rather than through online crowdfunding, cryptocurrency transactions, or other means.[69]

For international jihadism and some other violent movements, the ability to build operations across national borders is crucial. Social media platforms bring together resources and personnel across borders, and in some cases platforms even enable cross-border chains of command.

Social Media Enables Violent Groups in Rakhine State

Rakhine State, Myanmar, has a long history of tension and violence between the Rakhine Buddhist community and Muslims living in the region, mostly known as Rohingya. Government pressure compelled several hundred thousand Rohingya to flee to Bangladesh over the decades of military rule in Myanmar. In 2012, communal violence broke out in parts of Rakhine State, and the Arakan Rohingya Salvation Army (ARSA) subsequently emerged. Their poorly armed foot soldiers first attacked government border and police posts in 2016, using messaging apps to organize internally and to gather information to prepare for assaults. These attacks were used to justify a series of vigorous military sweeps against the Rohingya community, in which soldiers and associated militia committed massacres, violence against women and girls, rape, and mass arson. Around 700,000 Rohingya fled to neighboring Bangladesh, joining others already displaced.[70]

Social media platforms provided several protagonists with new tools that significantly shaped these events. The International Crisis Group reports that social media was critical to the structure and functioning of ARSA, including its creation by Rohingya diaspora activists with connections in Bangladesh, Pakistan, and possibly India.[71] Its governing committee of around 20 members was located overseas. ARSA leaders returned to the area to establish the organization and carry out operations. They also sought religious legitimacy, which they received through messages from foreign clerics—which, again, was only feasible through online channels, and which helped persuade local religious leaders to express support.[72]

Members of Rohingya communities in northern Rakhine were heavy users of WhatsApp, through which they maintained contact with family members abroad. Local WhatsApp groups shared social and security updates, often through voice recordings due to the lack of a written Rohingya language. Viber and WeChat were also in use, and all three platforms served ARSA's communications and planning. At the time of their border attacks, ARSA was unknown outside the immediate community; their first public statement was a video released on YouTube on October 14, 2016, with subsequent videos showing ongoing actions and stating their demands. These were crucial to establishing the group's identity and presence.

Despite the current lack of data, it is reasonable to surmise that social media connections can be key to the development of some mass movements, especially as they enable much faster membership growth with little or no "friction" between the impulse and the ability to join. Social media and messaging apps have become fully integrated into many violent extremist groups, and their use has largely evolved with the groups themselves. ARSA, a relatively new group, used messaging apps right from the start as essential tools for building their local networks. Without messaging apps, the group might not have been able to establish itself at all.[73]

Over the same period in which ARSA appeared, a much better equipped and trained ethnic Rakhine armed organization, the Arakan Army, also emerged. Growing in strength over time, the Arakan Army has used online media to gain support both within Rakhine State and elsewhere. Willing and able to conduct considerably more violent and damaging attacks than ARSA, the Arakan Army was by 2019 engaged in open warfare with the Myanmar government's armed forces across large swathes of territory.

The Online Terrain

In the same way that the physical terrain of a battlefield affects how a conflict is organized and carried out, the characteristics of social media platforms can affect or amplify the capabilities of conflict actors. Some of these characteristics result from algorithmic and interface design choices made by the developers of online platforms. For example, social networks that group people into like-minded communities may foster polarization as an accidental consequence of their design. In other cases, violent groups proactively choose their online terrain to fit their particular operational needs. Finally, local parameters of language, culture, and connectivity add another constraining factor to the ability of violent groups to operate online.

Shifting properties of social networks

Social network platforms that feature algorithm-driven recommendations may create "echo chambers" by emphasizing certain limited aspects of community identity. Social media platforms can make distant connections immediate—across the nation or the world—creating enormous, self-reinforcing communities based on a singular identity that excludes other affiliations and complexities. This phenomenon can have specific and profound implications in locations of conflict.

In areas with a plural or mixed population, neighborhoods are sometimes divided according to religion or ethnicity. Online echo chambers can exacerbate this dramatically: it is entirely possible that a person will see and hear more online from a member of his religious or ethnic in-group in another town or country than he will from a close neighbor who happens to be from a different group.[74] A narrative, opinion, or interpretation of an event that denigrates an out-group online will not encounter the voices or perspective of the targeted group; criticism intensifies, drawing confirmation and validation from participants in the in-group.

Facebook, by far the most-used platform, tends to generate the biggest effects and attract the most attention from those who aim to exacerbate division and prime the online space for conflict. Facebook's algorithms prioritize material that generates high engagement, including that which creates the most outrage. Motivated users can game this system by creating fake accounts and using built-in marketing tools to maximize their posts' reach—at least until they are detected by the company's screening systems. Inadvertently, then, a platform's features and algorithms can both prepare the ground for violent actors and give them the tools to further a conflict agenda.

Myanmar's Social Divisions and Facebook Echo Chambers

Divisions and clashes between Myanmar's Buddhist and Muslim communities have a long history dating back at least to colonial times—as do local histories of extended coexistence and collaboration.[75] Now that around 16 million citizens, or almost a third of the population, are on Facebook, these clashes are also playing out online.

Myanmar Facebook pages often organize along ethnic and religious lines that separate Buddhists, Muslims, and Christians, who might live in close proximity to each other in real life.[76] For example, when Buddhist or Muslim pages shared images of their own communities aiding flood victims in 2015, they commonly drew comments that "the other community" lacked the same virtue and did not help the needy. Muslim pages celebrated the Eid al-Adha festival by emphasizing values of charity and sharing with neighbors, while Buddhist pages noted the holiday with commentary condemning the brutality and cruelty to animals slaughtered for the ritual feast.[77]

Facebook's arrival in Myanmar coincided with the country's sudden opening to outside influences, and the platform quickly gained users by partnering with telecommunications operators to provide access with zero data costs.[78] Facebook also offered a Burmese language interface earlier than many other platforms, and it became the *only* way large numbers of first-time users in rural areas accessed the internet, supplanting traditional news outlets as the primary source of information.[79] The content that these users saw often existed within echo chambers that amplified existing prejudices and elided nuance. For instance, graphic depictions of violence in the Central African Republic were shared by Myanmar-based pages that misattributed the violence to Muslims, when in fact Muslims had been the victims.[80]

The 2018 UN Fact-Finding Mission on Myanmar systematically described how information on Facebook had contributed to violence.[81] In response, Facebook admitted its failure to act sooner, while noting that "Facebook alone cannot bring about the broad changes needed to address the human rights situation in Myanmar."[82] Facebook has since banned prominent instigators of violence and invested in local counterspeech programs and monitoring.

Choosing the right platform

Choosing the right platform can give conflict actors new capabilities. For actors in conflict with the state and therefore vulnerable to surveillance, the security of encrypted communications is vital. The ability to build networks across international borders and engage with large numbers of dispersed supporters who can provide skills or resources is also valuable.

Different platforms bring together different audiences, who can then be targeted for specific reasons. For example, in most locations across Asia, Facebook is the biggest platform by far, with 641 million total users across the region as of December 2019, which makes it the platform of choice to influence local or national public opinion. Representatives of international or global institutions—including

journalists, rights advocates, UN officials, and others—are more likely to be found on Twitter. This is therefore the platform to use to target global audiences, and evidence suggests this has happened in Myanmar and Indonesia, among other places.[83]

ISIS Uses Telegram for Hidden Communication

Jihadist networks have used different platforms at different junctures, depending on their features. Several leaders of the Marawi siege in Mindanao had first made contact with ISIS leaders in Syria via Facebook, a relatively open platform, before shifting to more secure options.[84] By the time of the siege itself, Telegram had become the prime communication channel for international jihadist extremists.[85]

Telegram was chosen over other products because it allows massive chat rooms—great for group discussions, community building, and broadcasting official messages—as well as encrypted chats—ideal for planning and conducting activities that must be hidden from the authorities or the general public. When one group was shut down, it was relatively easy to create a new group and continue broadcasting. While Telegram is a niche product compared to other messaging applications, ISIS could count on their supporters to reproduce public messages on other platforms for greater visibility.

Local parameters

Social media gets its power from network effects, the ability to connect as large an audience as possible in as many ways as possible. It allows for otherwise separate groups to communicate and identify with one another—across town, across a nation, or internationally. Both technical and human factors affect the scale of this interconnectivity, and therefore the impact of social media on a conflict.

Infrastructure is an overarching factor: audiences without internet access cannot promote their conflict to others, nor be so easily drawn into a conflict prosecuted by someone else. Poor or uneven connectivity may allow some sharing but reduces the speed and extent of engagement. Data costs can limit online communication or drive users to the cheapest platforms. As already noted, Facebook comes with free data in many Southeast Asian countries, so it becomes the platform of choice, while substantial data costs make video-streaming platforms much less popular. Given data costs and bandwidth constraints, some communities simply won't use platforms where certain violent actors are most engaged, and therefore won't be drawn into that world.

Language is also a significant parameter. Where communities involved in a conflict share a predominantly local language, tightly connected local networks can emerge, but shared content and collaboration with violent actors elsewhere is likely to be reduced or slowed. Some local languages, especially but not only those with a unique script, are not included in commonly used font packs for computers and smartphones,[86] forcing users to communicate in languages other than their native

tongue, to find cumbersome work-arounds (such as transliteration or communicating by voice recording) or simply to stay offline.

Limits of Connectivity and Language in Mindanao and Indonesia

The protracted conflict in Mindanao, in the southern Philippines, has been driven primarily by local and regional grievances, many of them among Mindanao-based actors as well as with the national government. Militant groups have largely been sustained by local clan networks.[87] Mindanao's online connectivity is limited, especially in poorer areas. A recent study found that Facebook is the most popular platform, largely because of its free data, while alternatives get little use.[88] Posts are predominantly in local languages, and Facebook networks in Mindanao are densely connected to one another but relatively unconnected outside the region.[89]

These factors helped to shape local online content during the occupation of Marawi City in 2017. This intense conflict changed the local dynamic by reframing local violence within ISIS's international extremist agenda. The occupiers were the Maute Group and the Basilan-based faction of Abu Sayyaf, both members of a local coalition that had pledged allegiance to ISIS.[90] In Marawi, the Maute Group's leadership networks provided the basis for operations on the ground.[91]

Social media platforms and encrypted messaging apps like Telegram were central to these efforts, both for public communication and operational purposes. Yet while the Marawi occupation drew on social media practices that had evolved within international jihadist networks, they made narrower use of them, mostly because of poor connectivity and local language barriers in Mindanao, demonstrating that social media use is still shaped by local conditions.

While the fight for the city dominated international jihadist discussion, the largest study of local social media activity supporting Mindanao's ISIS coalition found no dedicated Facebook pages for this purpose. Individual posts supporting ISIS on other pages were haphazard, with no evidence of a coherent strategy among local content producers.[92]

Notably, social media use by extremist jihadist groups in Indonesia spread faster and more widely across multiple platforms, including Facebook, Telegram, WhatsApp, and more recently Instagram.[93] Networks extend across the country and the diaspora, drawing on a large population, high levels of connectivity, a common language, and a long history of online and social media use across multiple platforms.[94]

Platform Companies' Responses to Violence

Every large social media company has had to formulate a response to the use of its platform for violence. These responses have varied widely and evolved rapidly depending on the company's home location, its core values, the changing political environment, and the outcomes of legal challenges. In general, major social media companies prefer to accept most content, while agreeing that certain material—and certain groups—should not be permitted on their platforms. These disallowed categories are generally outlined in terms of service or community guidelines and enforced through content moderation techniques that rely variously on voluntary action by users, algorithmic detection, and "trust and safety" teams that handle more nuanced cases. This section briefly outlines social media companies' decision-making about the prevention of violence and conflict, the ways they have begun working with other counterviolence stakeholders, and the limitations they face.

Legal frameworks

Limiting the use of online space for propagating violence in Southeast Asia depends in part on actions in the United States, where many influential tech companies are based. The freedom that online platforms enjoy to host user-generated content and to block or restrict access to that content is established in United States law by Section 230 of the Communications Decency Act of 1996.[95] Section 230 legally distinguishes "publishers," who publish their own content, from "interactive computer services," which provide a platform for the publication of content by third parties, such as YouTube videos, Twitter posts, and the comments sections of personal blogs. The law provides safe harbor to interactive computer services, protecting them from liability either for user-published content on their platform or for actions they take "in good faith" to remove or restrict objectionable content. These two provisions have been bedrock principles enabling free speech and debate to flourish on the 21st century internet.

When Section 230 was made law, the internet looked very different than it does today, and the influence of online platforms in actively curating the content that users see and in swaying mainstream public debate was not a major cause for concern. Now, the applicability of Section 230 to modern social media platforms is being tested, in part due to the use of these platforms by avowedly violent actors, including terrorists and extremists, and the perception that the platforms themselves are enablers.

In a 2019 case decided by the U.S. Court of Appeals for the Second Circuit, a group of Americans who were victims of a Hamas attack in Israel argued that Facebook had "unlawfully assisted" a terrorist organization when Hamas entities or supporters were able to post messages on their Facebook pages encouraging attacks in Israel, which the attackers allegedly read and acted upon, and when Facebook "connected" members of the Hamas network to each other via recommendation algorithms. While the court sided with Facebook, citing Section 230, a dissent by one of the judges highlighted the degree to which social media platforms have strayed from basic assumptions of the law—for example, by

deploying recommendation algorithms that generate real interpersonal networks, including networks explicitly committed to violence or conflict.[96]

Should the U.S. Congress choose to revisit Section 230 and update it—a not unlikely scenario given current political discussions—the legal requirements for online platforms to deal with violent and extremist groups could become more stringent worldwide. Already, other countries have begun introducing "intermediary liability" policies that require platforms to take some responsibility for posts by users.[97] Meanwhile, global cybercrime legislation has increasingly moved toward criminalization of dangerous or indecent online speech, leading to concerns that these laws will also curb peaceful free expression on the internet.

Moderation techniques

Even without strict legal requirements, however, high-profile online platforms are feeling pressure to limit the visibility of objectionable content and users, including conflict actors and their supporters. Using the legal cover provided by Section 230 and similar laws, most platforms have attempted to do this through self-regulation. They rely on a combination of algorithms that detect objectionable content and coordinated behavior, human moderators who review content and user profiles that may violate platform rules, and, in some cases, tweaks to their core product to allow users more freedom to understand, contextualize, and report content or users that may be fomenting violence or otherwise breaking the platform's rules.

Facebook's Moderation Efforts in Myanmar

The history of Facebook's content moderation policies in Myanmar offers a valuable case study in how a platform's internal systems can struggle in conflict environments. In 2013 and 2014, the presence of anti-Muslim hate speech on the platform was already apparent to local and international observers, and the links between online speech and real-world violence were underscored by the July 2014 mob violence in Mandalay, where false and incendiary accusations were spread by extremist leaders on Facebook.[98] Local leaders attempted to get these posts removed, but they could not keep pace with the rapid spread and had no direct channel to policy managers at the company. Facebook responded to these events by promising to improve its existing systems to better serve the Myanmar audience; for instance, by translating their community standards rules into Burmese.[99]

Facebook's responses in 2014 were broadly in line with its philosophy at the time, which sought to maximize access and use of the platform while minimizing censorship. But compared to larger markets like the United States, Facebook's moderation activities in Myanmar were not as sophisticated, effective, or well resourced. Facebook product and policy managers acknowledged these limitations and highlighted some specific ways that their moderation systems were failing in Myanmar: technical difficulties handling unique local fonts and scripts, which hamstrung their automated detection systems; lower usage of

reporting tools by Myanmar users, possibly because the reporting interface had not been fully translated into Burmese; and not enough human moderators to keep up with the waves of content generated by Myanmar users.[100]

Since 2018, Facebook has taken a more proactive approach to the use of its product for violence. It has hired thousands more moderators to review borderline material; created new teams spanning product, policy, and operations departments to study conflict and violence; and invested heavily in improving algorithms that help it identify hate speech before users report it.[101] In Myanmar, it specifically banned accounts connected to violent groups and hired 100 moderators who speak Burmese or local minority languages.[102] Bans of prominent and government-backed groups are now disclosed regularly in global transparency reports and public data releases.[103]

Each element of the moderation formula has inherent challenges. No platform wants to make their algorithms completely public, for fear of manipulation or criticism, and algorithmic errors or biases can result in overcorrections that limit the speech of vulnerable groups.[104] Engineering changes (such as adjustments to recommendation algorithms, forwarding limits, or other changes to a platform's underlying code), while critically important, are very expensive and can usually be circumvented by motivated users.

Social Media Company Crackdowns and Extremist Contingency Plans

Efforts to crack down on extremists' online communication include high-tech measures such as stopping the use of steganography (meaning techniques to conceal sensitive files within other files). Extremists' social media use has evolved in response to surveillance and crackdown risks, although it often involves low-tech steps that evade high-tech prevention, such as storing files on multiple platforms and avoiding algorithmic keyword searches by substituting terms (using *1515* instead of *ISIS, for example)*.

Media companies also monitor for bots, suspicious accounts, or infiltrated groups. Once infiltrated, even moderated groups may be flooded with pro-ISIS material. After the original members leave, these groups can operate for some time without being detected or banned. These strategies rely on personal networks, which allow connected individuals and groups to relocate when social media companies take down their accounts.

Twitter's mass takedown of extremist accounts was one reason why ISIS and its supporters switched to Telegram. Since Telegram launched large-scale takedowns in November 2019, ISIS members have been relocating globally to TamTam, Hoop, and other less well-known channels.[105] These platforms may not offer the same resources as Telegram, and reconstituting online groups takes time, but the effectiveness of one mass deplatforming will depend on how effective the next platform proves to be and the speed of its response in turn.

Collaboration with civil society

Faced with the complex challenges of violence, conflict, and abuse—along with tremendous public scrutiny—larger online platforms have signaled an interest in going beyond insular self-regulation and working more collaboratively with civil society to understand and address their users' behavior. New data-sharing agreements and public "transparency reports" have been important first steps, though the slow progress and limited scope of the disclosures have frustrated some researchers.[106]

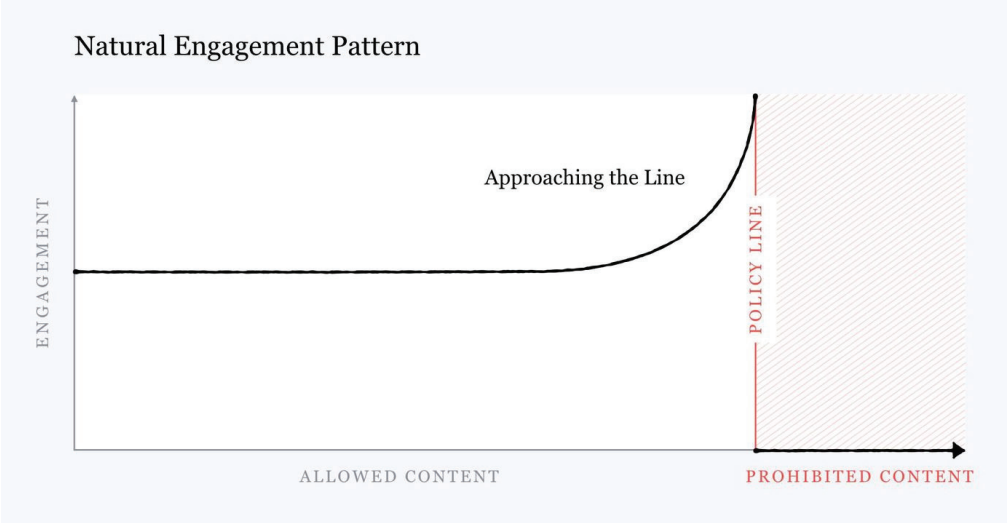
In 2019, Facebook announced plans (and $130 million in funding) for a more significant effort: an independent "oversight board" comprising 40 human-rights advocates, journalists, and other experts who will decide thorny content policy cases with real-world implications.[107] Cases will be chosen for their societal significance and the difficulty of the decision itself, and may provide a way for Facebook to avoid a repeat of its role in Myanmar, where it was directly blamed for enabling communal violence. Critics have complained that the board generates good publicity by taking on a limited set of high-profile cases, but has little effect on internal policies governing community standards, context-relevant moderation, and other issues. Another point of criticism is that the board's jurisdiction extends only to content removal decisions referred to it by Facebook itself or through an appeals process, and not to the many other decisions that the company makes about policy and enforcement.[108] Nonetheless, the board represents a step toward greater collaboration between online platforms and civil society.

Other efforts by platforms to mitigate conflict and violence center around the promotion of "counterspeech" rather than heavier moderation. The idea of counterspeech is essentially that the remedy for false or negative speech is *more true and positive speech*, rather than limits on expression, and it has almost a century of support in U.S. jurisprudence.[109] When called before U.S. lawmakers in 2018, representatives from Facebook, Google, and Twitter highlighted counterspeech efforts that they have undertaken to mitigate the influence of extremism, with examples ranging from support and training for civil society groups to "empower positive and moderate voices" to the use of targeted advertising tools to serve counterextremist content to people who might be becoming radicalized.[110]

At least eleven platform companies have also joined an industry-led initiative called the Global Internet Forum to Counter Terrorism (GIFCT), which aims to "significantly disrupt terrorist exploitation of the internet" through "support for expanding the capacity of civil society organizations to challenge violent extremism," as well as the creation of content standards, data transparency, and development of shared technical solutions. The GIFCT has evolved significantly since it was established in 2017, and it now holds regular consultations with civil society groups and releases transparency reports. But the limited influence of civil society representatives on the group's activities, and its limited track record to date, means it is hard to evaluate the group's effectiveness.[111]

Future issues and limitations

In a 2018 post announcing Facebook's plan for content moderation, Mark Zuckerberg presented an image of what he called the "natural engagement pattern" of social media users:[112]



According to this graph, the closer a piece of content comes to prohibited categories, the more people will be inclined to engage with it (e.g., by reading, liking, sharing, or commenting). Facebook's platform, like its competitors, uses engagement as a core metric of success: advertisers pay for it, page and group administrators attempt to optimize for it, and individual users feel rewarded by it. At the same time, if this graph were correct—and without access to Facebook data, it is hard to say for sure it is—it suggests that unscrupulous users could manufacture online engagement by sharing content that approaches the margin of permissibility, including violent and extreme materials. In Zuckerberg's 2018 announcement, he claimed that Facebook would adjust its algorithms to "flip" this engagement curve, limiting the spread of "borderline" content and punishing users who consistently approach the policy line. But two years later, it is not clear to outside observers just how effective these changes were, how regulation could be applied to hold companies accountable to such promises, or how much a platform can even do to push against the purportedly "natural" behavior of users.

A somewhat clearer target for platform companies is the coordinated abuse of their products by organized groups, such as state-backed disinformation operations that use targeted advertising and armies of fake accounts to steer online attention and debate toward particular topics. Platforms have responded by ramping up "bot" detection, requiring transparency from political accounts and ad buyers, and making public disclosures when they catch "coordinated inauthentic behavior." In 2019, Twitter and Facebook disclosed that they had detected and blocked operations originating in Bangladesh, China, India, Indonesia, Myanmar, Pakistan, the Philippines, Thailand, and Vietnam, as well as many more in other regions outside Asia. In the coming years, both state and nonstate actors will continue to challenge the abilities of online platforms and independent monitors to detect their increasingly sophisticated campaigns. In the absence of clear regulation or external oversight, the willingness of platform companies to challenge such abuse will be tested, especially when it is perpetrated by powerful incumbents.


NEW CASE
KNOW YOUR FLOW
23
Galaxy

Bringing Peacebuilding and Violence Reduction Online

Many organizations seek to limit the use of online platforms for violent objectives, or to use social media to promote tolerance and nonviolence instead. This section offers a brief overview of four different approaches, encompassing initiatives carried out by NGOs, civil society, and government security services. It draws particularly on examples from Myanmar, where many civil society initiatives are underway.

Building better understanding of online conflict dynamics

Developing a detailed understanding of the multiple and complex roles that social media plays in conflict dynamics is a first and necessary step for mitigation efforts. For example, Myanmar has seen several research efforts that provide detailed insight into social media and conflict dynamics, including how these connect to and play out in the physical world across different locations. While many gaps inevitably remain, these examples show the value of different types of inquiry.

The Myanmar Media and Society Research Project

The Myanmar Media And Society (M.MAS) research project began by mapping how a chauvinist narrative of Islam as a threat was generated by majoritarian Buddhist interests.[113] The project identified the arguments that were deployed to support this narrative. The research questions the strong emphasis that those working to mitigate conflict place on countering "rumors" and "hate speech," and argues that a narrative that makes "others" of Muslims does not necessarily need to resort to overt hate or vilification.

The project then explored in detail those locations that remained peaceful even though the conditions for an outbreak of violence were present, including rumors and outright calls for attacks.[114] It asked what prevented violence, especially considering how local interventions were carried out. The research highlighted many local differences along with some common features. In particular, individuals who had built up trust and credibility across communities and with authorities, and acted to defuse potential conflict, were critical to maintaining peace. Although several of these individual "peacemakers" were able to cross religious lines, they were often not themselves religious figures, but rather land activists, advocates for the disadvantaged, community elders, or other respected local figures.

The arguments that these peacemakers used ranged from the damage that violence would inflict on all parties, to alternative pathways to resolution, to invoking peaceful coexistence

in the past, to highlighting the role of political manipulation in generating the tension. Cases included examples of online incitement and others examples in which this was not a significant factor; that is, it located the role of social media within local conflict dynamics more broadly rather than treating social media as the sole or primary factor, thereby exposing where social media platforms were indeed a driver of conflict as well as where they were not.

Online Use and Youth Culture: Avoid the Stereotypes

Research by Phandeeyar, and by Save the Children and the University of Sydney, explores the attitudes and understandings of young users of online platforms.[115] Researchers analyzed how young users understand and engage with hate speech, and which sources they trust. The resulting studies criticized widely held assumptions that Myanmar users were universally digitally illiterate and passive consumers of online information. Instead, the studies found that, despite a number of digital skills deficits, engagement with online content was also active and framed by local attitudes (such as views towards privacy that differ from those in the West) and historical experience, including practices for checking rumors.[116]

Online counter-narratives

Some conflict mitigation focuses on developing online peace messaging or counter-narratives, or on opening or maintaining space for alternative and moderate views. Related approaches encompass sustained, proactive initiatives as well as short-term, reactive campaigns that respond to specific events.

Combating Hate Speech Offline and Online

Myanmar's Panzagar is an anti-hate-speech campaign that started offline at the 2014 ASEAN People's Forum in Yangon, with participants holding signs saying "do not spread hatred" as a counter to extreme nationalist groups.[117] This offline effort led to the development of Facebook stickers with amusing characters and positive messages. The campaign generated 2.7 million downloads in the 12 months ending in February 2016, and the stickers were used in messages 12.9 million times.[118] Several offline activities, including pamphlet distribution, regional trips, and launch events were also incorporated. The campaign's strengths, according to observers, included extensive collaboration among civil society actors that combined creative ideas and the skills needed to bring them to fruition.[119]

The Panzagar initiative, unlike many others, generated sustained action. It led to the foundation of MIDO, one of Myanmar's highest-profile and most active organizations in conflict mitigation, online initiatives, and community-level peacebuilding efforts. Other activities include consulting with communities on memories of peaceful coexistence to

create "a log book of positive messages that can then be communicated through various channels."[120]

Fostering a Fact-Checking Community

Phandeeyar's Tech for Peace team established The Burma Monitor in 2017, with a "Think Before You Trust" campaign. The team uploaded examples of fake online news, which were debunked with fact checking or photo analysis. Lessons were shared along with advice on how to respond to fake news. The campaign developed an online community that began to submit their own examples. It was particularly successful in engaging young people: 94 percent of its Facebook fans were between 18 and 34 years old, and the page gained 65,000 likes.[121]

A recent global review of literature in this field by UNESCO identifies several practices, issues, and questions, but also highlights a lack of empirical research into the effectiveness of counter-narrative approaches. Objective impact evaluations have typically been inadequately funded, while research and practice have rarely drawn on other relevant disciplines such as psychology or the history of propaganda. Counter-narrative campaigns may inadvertently support the efforts of the violent actors they mean to oppose by "granting [radicalizers] power over setting the agenda and the narratives and helping them monopolize attention to their issues to the detriment of other realities."[122] The literature review also notes that " there is a big question as to whether findings in liberal environments have any resonance in more closed ones, and vice versa…. Partly for this reason, it can also be questioned whether a search for general causal effects in this area can ever be definitively resolved."

Local peacebuilding

Local peacebuilding assumes many forms. It often draws on traditional forms of conflict resolution or seeks to empower people who seek to broker an end to local disputes. Some initiatives support the roles of existing authorities such as village elders and religious leaders, while others look to strengthen the role of women, youth, or the community at large.

Local peacebuilding efforts have begun to incorporate online strategies, usually by responding at the community level to online rumors or incitement with alternative online content and with interventions in the physical world. Such efforts may seek to identify and engage individuals perceived to be at risk of participating in extremist actions, or they may aim to address community-wide concerns over tensions and violence. Local peacebuilding is pursued in both cities and rural areas, and targets specific institutions such as universities and sites of worship as well as urban neighborhoods or villages.

Locally Tailored, Community Conflict Mitigation in Lashio and Mandalay

Community Information Management to Reduce Inter-Communal Violence, an innovative peacebuilding project led by MIDO and Search for Common Ground, began in two cities in Myanmar in 2015. The work explicitly builds on prior research and on both organizations' familiarity with the two locations.[123] Initial steps involved setting up a community information management committee of influential local residents. The makeup of the committee reflects local context, with several religious figures included in Lashio, but with participants largely identified by vocation in Mandalay, reflecting the specific nature of conflict in each place. Committee members received training in understanding rumors, their manipulation, and how this can result in violence; and in how to respond, especially through collaboration with other stakeholders.

An inception study and later evaluations included cases in which rumors or misinformation had caused heightened tension but had then been countered, including false online reports about the killing of a Buddhist monk by a Muslim man in a nearby village. Project participants shared information via Facebook, through real-world community outreach, and through selected media outlets that boosted the credibility of the mitigation effort. The role of physical locations—tea shops and beer halls—in generating and disseminating rumors was explicitly acknowledged.[124]

Policing of social media

Government security agencies seek to mitigate the effect of online platforms in conflict, typically targeting designated terrorist groups and less frequently others such as majoritarian chauvinist movements. Measures include both surveillance and intelligence gathering for eventual arrests or other field actions, and account or website shutdowns. Across this complex and challenging field, governments have been frustrated by the weak response of social media companies to their calls for action; sometimes companies only react when a government has regulated or restricted their operations.

While policing violent actors online is vital, the same measures are often turned against legitimate social protest, whether through mission creep or by design. Human- and women's-rights advocates argue that governments should base their approach on respect for the rights of citizens and the rule of law, rather than securitizing or militarizing civic digital space. Government engagement may also be compromised when individuals or agencies within the state are themselves conflict protagonists, a common scenario when conflict pits the state against armed groups.

Government Efforts to Reduce Online Extremist Activity in Indonesia

The Indonesian government's practice of taking down extremist websites was rendered moot by the shift to social media—in particular to encrypted chat, with Telegram quickly becoming the most popular app.[125] Telegram did not immediately respond to government requests for action—Facebook, Google, Twitter, and others responded with more alacrity—until the government blocked the platform's web access in 2017. Since then, the company has worked more closely with the government, and several larger and more open Telegram groups have since been identified and taken down, although encrypted private chats are still impenetrable to authorities. Other initiatives include Google's partnership with Indonesia's Ministry of Communications to develop "trusted flaggers" to identify content to be taken down. From the companies' perspective, distinguishing legitimate political expression from hate speech or criminal incitement remains a challenge, even though some progress has been made. The working relationship between a government and social media companies is key to the effectiveness of these measures. The Philippines government, for example, has failed to develop these relationships and so is less nimble in this space.[126]

Indonesia's Ministry of Communications, following a prison riot by extremists, claimed that its Cyber Drone 9 system, which uses machine learning, public reports, and human verification, had identified 22,000 pieces of "radical" content and blocked 4,000 of them.[127] The ITE Law provides the legal basis for suspending content that incites, and for any charges that may follow; however, the law also includes vague definitions of defamation that could be exploited arbitrarily.[128]

Nongovernmental Prevention of Violent Extremism in Malaysia

The religiously grounded work of Angkatan Belia Islam Malaysia (ABIM), or the Malaysian Islamic Youth Movement, is an example of a prevention initiative by a nongovernmental group.[129] To identify extremist sympathizers, ABIM staff track their movements using social media, working to understand patterns of behavior and identify points where intervention can disrupt the radicalization process. After identifying individuals who are in the process of radicalization, ABIM works with them to promote a better understanding of positive religious teaching and outlooks. While this approach has not been systematically evaluated, ABIM staff say they have repeatedly seen changes in extremist sympathizers after interacting with them over many years. According to interviewees, changes in online behavior and language are especially noticeable.

To date, the program has reached 60,000 individuals. Other ABIM initiatives include the Malaysian Lifeline for Syria, a humanitarian aid effort, supported by the Global Peace Mission, to provide clothing and access to clean water, schools, and education to displaced persons in Syria. It has also supported public events promoting tolerance, such as the mass Solidarity for Peace March in 2019, and they have backed programs promoting mosques to the public as open centers for dialogue and learning.

In Thailand's far south, a local civil society group called The Motive generated a series of short videos for social media to raise awareness about Covid-19 and counter misinformation and hate speech directed towards their community.

Recommendations for Action

Non-governmental groups, academics, public or private funding organizations, and platform companies should strengthen independent monitoring and evidence-based research on the online dimensions of violence.

Empirical data and evidence are critical for the development of responses to the trends and incidents identified in this report. Currently, for violence occurring in the physical world, indicators managed by independent groups like the Uppsala Conflict Data Program (UCDP), the Armed Conflict Location and Events Data Project (ACLED), and locally managed programs like Thailand's Deep South Watch and Myanmar's Township Conflict Monitoring System provide important data on violent incidents. They inform early warning systems, interventions by government and civil society, and peacebuilding and reconciliation efforts.

Such groups are increasingly engaging with online platforms, both as a source of evidence for physical events and as a field to monitor in its own right. But there are few independent databases on the online dimensions of conflict, and the parameters are not as well defined, making it hard to assess the overall prevalence and severity of online activities that promote violence.

For Southeast Asia specifically, while a number of studies examine social media's impact on some violent conflicts in the region, many more conflicts have been only partially examined, if at all. The issue is significantly understudied in Southeast Asia compared to the Middle East and the West. Documentation of specific mitigation efforts in Southeast Asia is also patchy. Given the increasing integration of platforms into conflict dynamics, these research gaps represent a serious obstacle to understanding and developing effective responses to violent conflicts across the region.

Globally, the greatest barrier to the creation of independent research and data-collection programs is the control of user data by platform companies themselves, which act both out of concern for user privacy and out of a desire to preserve their competitive advantage. As a result, platform companies are far more able to analyze user behavior and its correlation with conflict and violence than independent researchers. While many large platforms have dramatically improved their transparency reporting, direct data-sharing partnerships between platforms and independent researchers or civil society groups is still limited and inconsistent.

Meanwhile, many of the largest U.S.-based platform companies are finding common ground, not with counterparts in civil society and government, but with each other. As a case in point, the Global Internet Forum to Counter Terrorism (GIFCT) was formed by several large platform companies to support joint research and collaboration on terrorist use of their platforms, emulating many of the successful techniques used to remove child sexual abuse materials from the internet. But while the software tools and databases used to find and remove this material are coordinated by independent child-safety nonprofits with oversight from law enforcement, the GIFCT is controlled by the technology companies themselves, with limited oversight or guidance from civil society.

As the links between online activity and offline violence grow closer and more obvious, and as social media companies continue their steps toward transparency and disclosure, it will be critical for academics, civil society organizations, and independent, nonpartisan research groups to lead the way in defining what shared responsibility for monitoring and data collection will look like, and how government oversight or regulation should take place. Progress in differential privacy techniques, which allow confidential data to be made available for analysis without causing harm to the individuals represented in a database, could rapidly accelerate this shift toward openness.[130]

Civil society groups and partners can develop agendas for coherent, evidence-based research into the use of online platforms in conflicts. As connectivity increases and sophisticated digital tools become more readily available, information on these evolving patterns of use need consistent monitoring and continuous updating. Civil society groups and academics can develop their own capabilities, understanding, and best practices by conducting open-source investigations and establishing centers for the rigorous study of online communities. Relevant models include the Berkman Klein Center at Harvard, the Stanford Internet Observatory, and notable research initiatives from Asian think tanks and NGOs like the Center for Policy Alternatives (in Sri Lanka) and the Institute for Policy Analysis of Conflict (in Indonesia).

Regional development actors and international donors can support this research, particularly in understudied areas of Southeast Asia. In addition to direct capacity-building support, donors can promote international linkages between civil society groups and researchers and support the sharing of relevant findings.

Action points

- Support in-depth research to study how groups are using online platforms to foment conflict in the region.
- Specialist research and monitoring groups should develop data-driven indicators of online aspects of conflict such as hate speech, harassment and bullying, online recruitment and advertising by known violent groups, and network polarization. Consistent monitoring of these indicators can help identify where hate speech or negative sentiments are on the rise. It can also be used to hold platform companies accountable by assigning clear and measurable targets to their conflict-prevention activities.
- Public, nongovernmental, and corporate funders should fund relevant research and training. Independent groups are often well positioned to conduct context-sensitive analysis and research—for instance, by accessing existing information and harnessing new technical tools—if they are offered resources and support.
- Develop international and regional linkages for civil society groups, researchers, and governments to share findings and draw lessons from other parts of the world where resources on conflict and online platforms are more extensive.
- Within the technology sector, change intra-industry counter-extremism databases and tools to allow participation and oversight by trusted, independent, civil society groups with subject-matter expertise in violence and conflict.

Governments, through consultation and with external support where needed, should develop balanced policy frameworks that promote inclusion and shared norms.

The growing significance of online platforms is causing governments around the world to reconsider platform liability and the policing of online speech. As the digitization of conflict continues, governments can do more to enforce platform companies' commitments to transparency reporting, proactive enforcement of community standards against violence and conflict, and support for civil society peacebuilding efforts. Many of the major platforms have already committed to these principles under the 2019 Christchurch Call to Eliminate Terrorist and Violent Extremist Content Online, a pledge made by governments and technology companies following the 2019 attacks on mosques in Christchurch, New Zealand, by an individual who was radicalized online and live-streamed the attacks on Facebook.

Governments can use the momentum of this pledge to strengthen communication with platform companies and, when new rules against online harm or intermediary liability are proposed, to consult widely in order to evaluate the real effects of policy changes on users, institutions, and digital entrepreneurs. These initiatives should include both large and small online platforms, recognizing that smaller and newer companies may not immediately have the data or financial resources to support content moderation at the scale of Facebook, Twitter, or Google.

International norms and standards are lagging behind emerging varieties of cyber-conflict. Malicious online behaviors can do real damage to civic institutions, yet they are hard to attribute to particular actors and do not fall within accepted definitions of warfare. For example, online influence operations are now widely reported around the world, but very few groups have the technical skills to trace and identify sources. Efforts by large online platforms to improve transparency are criticized as slow and incomplete, and definitions of popular concepts like "bots" and "deepfakes" are vague and often disputed.[131] Catching up to this problem will require regulation which consolidates norms and standards, and is based on more consistent dialogue, research, and engagement between public-sector, private-sector, and civil society stakeholders.

Action points

- Inter-governmental bodies, governments, NGOs, and the private sector should hold dialogues on the governance of online platforms, with a special focus on the intersections between violent conflict and policy issues like encryption, cybersecurity, and cybercrime.
- Develop clear standards and practices for countering online activity that promotes conflict. In particular, consider efforts by political interests or conflict actors to game or manipulate social media.
- Governments making policy to address online aspects of conflict should consult widely rather than deferring to security agencies that operate with limited oversight or accountability.
- Donors, regional and global intergovernmental bodies, and platform companies should work with governments to build expertise in regulatory frameworks that adhere to international principles of human rights.

Platform companies should refocus platform policy development and monitoring to support local leadership and engagement.

Policies for moderating online content and enforcing terms of service have a relatively short history, and until very recently they have been treated as internal matters for tech companies. Now that online platforms are such an integral part of public life, these policymaking practices must be reformed to create greater transparency, consumer safety, and trust. This need is especially acute in conflict-affected regions and in areas where automated systems cannot process local languages.

Though circumstances vary, online platform companies typically assign sensitive policy issues in smaller or less strategically important markets to regional staff, who may have little to no knowledge of local context or languages. This approach is inadequate in much of Southeast Asia, which has many different types of conflict, thousands of local languages, and a multiplicity of smaller armed groups in peripheral areas. Addressing this context requires an approach that is sensitive to local conditions and can be adjusted through audits and local dialogue.

With better technology, expanded policy-review teams, and more experience in the field, large platform companies have improved policy enforcement and the detection of abuse. These are laudable achievements. But to maintain peace and safety both online and in the real world, platforms will need to work in a different way. Besides investing in centralized algorithms and decision-making procedures, platforms should establish dialogue and auditing mechanisms that give a voice to local stakeholders in the development and enforcement of platform policies in conflict-affected areas.

Action points

- Recognize that responding to conflict in Southeast Asia and further afield normally requires an understanding of local conditions. Subnational perspectives must be included in platform policymaking by hiring people with practical, on-the-ground expertise and adapting to the use of minority languages.
- Conduct impact assessments for major new features and products that include analysis of their potential effects in conflict-ridden areas. Current practice largely confines third-party input to the policy-review process; this should be extended to product changes as well as policy adjustments, to enable both product and policy teams to gauge the impact of their work, not just on growth metrics, but on actual outcomes among vulnerable populations.
- Platform companies, donors, and others should back training for local reporting of hate speech, misinformation, and disinformation, ensuring that gendered aspects of the conflict-technology nexus are integrated into such support.

All parties should support innovation in peacebuilding and conflict-prevention practices.

An asymmetric capacity gap has emerged between violent actors who use platforms to promote conflict and peaceful groups who often lack the tools and capabilities to respond. This asymmetry does not need to be permanent: positive messaging does spread widely on social media, even in deeply conflict-affected areas, and local peacebuilding organizations have had success promoting such messages online, especially when these online actions are combined with real-world activities.

Support from donors, platform companies, and other actors can help these local peacebuilding groups take on larger and more effective roles. For example, local or national groups can participate in fact-checking and awareness raising to reduce the spread of incendiary misinformation, disinformation, and online harassment. These groups can also integrate discussion of good digital citizenship into community dialogues. Finally, because technological and social issues increasingly overlap, peacebuilding organizations should start discussing technology policy, especially cybersecurity, consumer safety, and policing online speech.

Dealing with online contributors to violence means navigating a new landscape with its own rules and policing mechanisms. Platform companies have acknowledged that reporting tools designed to flag false and harmful content are used infrequently, particularly by less digitally literate users. Users who do report violent or hateful content are often discouraged when no action is taken. Platform companies must design reporting systems that are effective in all environments with the same rigor they bring to other design challenges. Local groups, NGOs, governments, and aid agencies can help by promoting public awareness of ways to track, report, and investigate online violence and hate. Voluntary hubs for tracking otherwise unreported trends and incidents and monitoring responses should also be considered.

Lastly, when peacebuilders and mediators bring their work online, they should seek out rigorous and ongoing evaluation to determine the best methods for this new space. As social media becomes an increasingly important factor in conflicts around the world, platforms also become key to conflict mitigation. Conflict-prevention mobilizers and peacebuilders in Southeast Asia have looked to harness the internet in many ways, from global advocacy to community networking. Efforts to create online narratives to counter those promoted by conflict protagonists have become increasingly sophisticated, but effective monitoring and evaluation are still limited. Beyond pilot programs and one-off projects, a rigorously measured online counterspeech campaign would be a major contribution to our understanding of online platforms and violence.

Action points

- Local, national, and international agencies, both governmental and nongovernmental, can improve the depth and detail of the evidence base for online conflict-mitigation techniques, including counterspeech. Effective monitoring and evaluation are essential to understand the problem and find solutions. Given the continual and rapid change in the landscape of online conflict, monitoring must be ongoing.
- Governments, researchers, and nongovernmental peacebuilding bodies must better

understand the effects of counter-narrative initiatives and adapt their approaches accordingly.

- Educators, civil society organizations, and development actors should improve digital media literacy at the grassroots and build local capacity to manage relevant online dimensions of conflict, including countering hate speech on social media, using existing reporting tools to flag violent content, and identifying manipulated media.
- Researchers, local activists, governments, and companies should act together to address areas of common interest, such as tackling hate speech or preventing recruitment into violent extremism. These stakeholders can establish shared online tools and databases that enable collective action by the region's peacebuilders.

Endnotes

1 The Asia Foundation (2017), *The State of Conflict and Violence in Asia (*Bangkok: The Asia Foundation), https://asiafoundation.org/publication/state-conflict-violence-asia/.

2 A growing number of governments sponsor or direct online influence operations, usually covertly. Many of these operations have pioneered new ways of hacking and manipulating social media.

3 The Asia Foundation and Rappler (2018), *Understanding Violent Extremism: Messaging and Recruitment Strategies on Social Media in the Philippines* (Philippines: The Asia Foundation and Rappler), https://asiafoundation.org/publication/understanding-violent-extremism-messaging-and-recruitment-strategies-on-social-media-in-the-philippines/; Institute for Policy Analysis of Conflict (2017a), *Marawi, the "East Asia Wilayah" and Indonesia,* IPAC Report No. 38 (Jakarta: IPAC), http://www.understandingconflict.org/en/conflict/read/61/Marawi-The-East-Asia-Wilayah-and-Indonesia; Institute for Policy Analysis of Conflict (2016) *Pro-ISIS Groups in Mindanao and their Links to Indonesia and Malaysia,* IPAC Report No. 33 (Jakarta: IPAC), http://www.understandingconflict.org/en/conflict/read/56/Pro-ISIS-Groups-in-Mindanao-and-Their-Links-to-Indonesia-and-Malaysia.

4 Samidh Chakrabarti and Rosa Birch (2019), "Understanding Social Media and Conflict," About Facebook website, June 20, about.fb.com/news/2019/06/social-media-and-conflict/.

5 Julia Carrie Wong (2019), "'Overreacting to Failure': Facebook's New Myanmar Strategy Baffles Local Activists," *The Guardian,* Feb. 7, www.theguardian.com/technology/2019/feb/07/facebook-myanmar-genocide-violence-hate-speech. Colin Lecher (2018), "Myanmar Activists Say Facebook's Plans to Stop Violent Speech Are 'Nowhere near Enough,'" *The Verge*, Apr. 9, www.theverge.com/2018/4/9/17216810/facebook-mark-zuckerberg-letter-myanmar-activists.

6 The Economist (2020), "The Taliban Are Joining Afghanistan's Fight against Covid-19," *The Economist,* May 9, www.economist.com/asia/2020/05/09/the-taliban-are-joining-afghanistans-fight-against-covid-19.

7 Jason Burke (2020), "Opportunity or Threat? How Islamic Extremists Are Reacting to Coronavirus," *The Guardian*, Apr. 16, www.theguardian.com/world/2020/apr/16/opportunity-or-threat-how-islamic-extremists-reacting-coronavirus.

8 Tashny Sukumaran (2020), "As Malaysia Battles Covid-19, Its Rohingya Refugees Face a Torrent of Hate," *South China Morning Post,* Apr. 28, www.scmp.com/week-asia/politics/article/3081958/malaysia-battles-coronavirus-its-rohingya-refugees-face-torrent.

9 Emma Goldrick (2020), "Cambodia's Hun Sen Regime Introduces Repressive Emergency Laws under Cover of Covid-19," *New Mandala,* Apr. 20, www.newmandala.org/cambodias-hun-sen-regime-introduces-repressive-emergency-laws-under-cover-of-covid-19/.

10 Adrian Wail Akhlas (2020), "Indonesian economy could shrink 3.9 percent if hit by second Covid-19 wave, OECD warns," *Jakarta Post*, June 11, https://www.thejakartapost.com/news/2020/06/11/indonesian-economy-could-shrink-3-9-percent-if-hit-by-second-covid-19-wave-oecd-warns.html.

11 Christchurch Call (2019), *The Christchurch Call to action: To eliminate terrorist and violent extremist content online*, https://www.christchurchcall.com/call.html.

12 Sarah Perez (2019), "Reddit's Monthly Active User Base Grew 30% to Reach 430M in 2019," *TechCrunch*, Dec. 4, techcrunch.com/2019/12/04/reddits-monthly-active-user-base-grew-30-to-reach-430m-in-2019/.

13 Kokil Jaidka et al. (2019), "Brevity Is the Soul of Twitter: The Constraint Affordance and Political Discussion," *Journal of Communication* 69 (4): 345–372, http://academic.oup.com/joc/article/69/4/345/5547032?guestAccessKey=54a38170-de3f-4437-bbdc-cb3aa1ba1b5e. Mariska Kleemans (2018), "Picture Perfect: The Direct Effect of Manipulated Instagram Photos on Body Image in Adolescent Girls," *Media Psychology* 21 (1): 93–110, www.tandfonline.com/doi/full/10.1080/15213269.2016.1257392.

14 Kevin Roose, (2019), "YouTube's Product Chief on Online Radicalization and Algorithmic Rabbit Holes," *New York Times*, Mar. 29, www.nytimes.com/2019/03/29/technology/youtube-online-extremism.html.

15 End-to-end encryption for two-party chat conversations is increasingly found on mainstream services such as WhatsApp, Viber, and LINE, either by default or as an optional setting, a major shift from a decade ago.

16 Christian Christensen (2011), "Twitter Revolutions? Addressing Social Media and Dissent," *Communication Review* 14 (3): 155–157, https://doi.org/10.1080/10714421.2011.597235; Zeynep Tufekci and Christopher Wilson (2012), "Social Media and the Decision to Participate in Political Protest: Observations from Tahrir Square," *Journal of Communication* 62 (2): 363–79, https://doi.org/10.1111/j.1460-2466.2012.01629.

17 Imran Awan (2017), Cyber-Extremism: Isis and the Power of Social Media, *Society 54*: 138–149, https://doi.org/10.1007/s12115-017-0114-0.

18 James P. Farwell (2014), "The Media Strategy of ISIS," *Survival* 56 (6): 49–55, https://doi.org/10.1080/00396338.2014.985436.

19 Brendan I. Koerner (2017), "Why ISIS Is Winning the Social Media War," *Wired*, May 1, www.wired.com/2016/03/isis-winning-social-media-war-heres-beat/.

20 Simone Molin Friis (2015), "'Beyond anything we have ever seen': beheading videos and the visibility of violence in the war against ISIS," *International Affairs* 91 (4): 725–746, https://doi.org/10.1111/1468-2346.12341.

21 Ali Fisher (2015), "Swarmcast: How Jihadist Networks Maintain a Persistent Online Presence," *Perspectives on Terrorism* 9 (3): 3–20, www.jstor.org/stable/26297378.

22 J.M. Berger (2016), *Nazis vs. ISIS on Twitter: A Comparative Study of White Nationalist and ISIS Online Social Media Networks*, Occasional Paper Series (George Washington University), https://extremism.gwu.edu/sites/g/files/zaxdzs2191/f/downloads/Nazis%20v.%20ISIS.pdf; Max Fisher (2019), "White Terrorism Shows 'Stunning' Parallels to Islamic State's Rise," *New York Times*, Aug. 5, www.nytimes.com/2019/08/05/world/americas/terrorism-white-nationalist-supremacy-isis.html; Emerson T. Brooking and P. W. Singer (2016), "War Goes Viral: How Social Media Is Being Weaponized Across the World," *Atlantic*, November, www.theatlantic.com/magazine/archive/2016/11/war-goes-viral/501125/.

23 Nava Nuraniyah (2019), *The Evolution of Online Violent Extremism in Indonesia and the Philippines,* Global Research Network on Terrorism and Technology Paper 5 (London: Royal United Services Institute for Defense and Security Studies), https://rusi.org/sites/default/files/20190711_grntt_paper_5.pdf.

24 Paul Mozur (2018), "A Genocide Incited on Facebook, with Posts from Myanmar's Military," *New York Times,* Oct.15, www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html.

25 Disinformation is false information deliberately produced or disseminated by parties who know it to be false. Misinformation is false information not known to be false.

26 International Crisis Group (2017), Buddhism and State Power in Myanmar, Asia Report No. 290 (Brussels: ICG), https://www.crisisgroup.org/asia/south-east-asia/myanmar/290-buddhism-and-state-power-myanmar; Matthew Schissler, Matthew J. Walton, and Phyu Phyu Thi (2015), Threat and virtuous defence: Listening to narratives of religious conflict in six Myanmar cities, (Myanmar Media and Society Project), https://www.sant.ox.ac.uk/sites/default/files/m.mas_working_paper_1.1_-_threat_and_virtuous_defence_-_july_2015.pdf.

27 ICG 2017 and Schissler et al. 2015 (note 26).

28 Zarni Mann (2015), "5 Accused of Sparking Mandalay Riots Sentenced to 21 Years Jail," Irrawaddy, Mar. 19, www.irrawaddy.com/news/burma/5-accused-of-sparking-mandalay-riots-sentenced-to-21-years-jail.html.

29 Thomas Fuller and Wai Moe (2014), "Buddhist-Muslim Mayhem Hits Myanmar's No. 2 City," New York Times, July 3, www.nytimes.com/2014/07/04/world/asia/buddhist-muslim-mayhem-hits-myanmars-no-2-city.html.

30 An accurate translation of Ma Ba Tha's full, original name is "Association for the Protection of Race and Religion."

31 For more detail on Ma Ba Tha itself, see ICG 2017 (note 26).

32 Wirathu's Facebook accounts and those of several others were eventually suspended by the company. See the *Platform Companies' Responses to Violence* section. Sources: Facebook website, "Update on Myanmar," August 15, 2018, https://about.fb.com/news/2018/08/update-on-myanmar/; Laignee Barron (2018), "Nationalist Monk Known as the 'Burmese bin Laden' Has Been Stopped From Spreading Hate on Facebook," Time, February 28, https://time.com/5178790/facebook-removes-wirathu/.

33 BBC (2018), "UN: Facebook has turned into a beast in Myanmar," BBC, March 13, https://www.bbc.com/news/technology-43385677.

34 Carmela Fonbuena (2017), "Marawi Battle Zone: Urban Warfare Challenges PH Military," Rappler, June 19, www.rappler.com/newsbreak/in-depth/173050-battle-zone-marawi-urban-warfare.

35 For greater detail, see, for example, Fermin Adriano and Thomas Parks (2013), The Contested Corners of Asia: The Case of Mindanao, Philippines (The Asia Foundation), https://asiafoundation.org/publication/the-case-of-mindanao-philippines/.

36 IPAC 2016 (note 3).

37 Larry Greenemeier (2018), "Social Media's Stepped-Up Crackdown on Terrorists Still Falls Short," Scientific American, July 24, www.scientificamerican.com/article/social-medias-stepped-up-crackdown-on-terrorists-still-falls-short/.

38 IPAC 2017a (note 3).

39 Charles Knight and Katja Theodorakis (2019), The Marawi Crisis—Urban Conflict and Information Operations (Barton, ACT: Australian Strategic Policy Institute), www.aspi.org.au/report/marawi-crisis-urban-conflict-and-information-operations.

40 IPAC 2017a (note 3).

41 Rukmini Callimachi (2016), "A News Agency With Scoops Directly From ISIS, and a Veneer of Objectivity," New York Times, Jan. 14, www.nytimes.com/2016/01/15/world/middleeast/a-news-agency-with-scoops-directly-from-isis-and-a-veneer-of-objectivity.html.

42 IPAC 2017a (note 3).

43 See, for example, Yasmeen Serhan (2019), "The Common Element Uniting Worldwide Protests," Atlantic, November 19, https://www.theatlantic.com/international/archive/2019/11/leaderless-protests-around-world/602194/.

44 Merlyna Lim (2013), "Many Clicks but Little Sticks: Social Media Activism in Indonesia," Journal of Contemporary Asia 43 (4): 636–657, https://www.tandfonline.com/doi/abs/10.1080/00472336.2013.769386.

45 San Yamin Aung (2020), "Myanmar Called to End Yearlong Internet Blackout in Volatile West," Irrawaddy, June 22, https://www.irrawaddy.com/news/burma/myanmar-called-to-end-yearlong-internet-blackout-in-volatile-west.html.

46 Rebecca Ratcliffe and Redwan Ahmed (2020), "Bangladesh urged to lift Rohingya internet ban as Covid-19 rumours swirl," Guardian, June 11, https://www.theguardian.com/world/2020/jun/11/internet-ban-sparks-covid-19-rumours-in-rohingya-camp.

47 Human Rights Watch (2020), "Myanmar: end unlawful internet restrictions," Human Rights Watch website, July 27, https://www.hrw.org/news/2020/07/27/myanmar-end-unlawful-internet-restrictions.

48 Dave Kendall (2019), "Dictating the internet," Bangkok Post, December 16, https://www.bangkokpost.com/thailand/special-reports/1816894/dictating-the-internet.

49 For example: Benjamin Strick (2019), Investigating Information Operations in West Papua: A Digital Forensic Case Study of Cross-Platform Network Analysis (Bellingcat), https://www.bellingcat.com/news/rest-of-world/2019/10/11/investigating-information-operations-in-west-papua-a-digital-forensic-case-study-of-cross-platform-network-analysis/.

50 Lorcan Lovett (2017), "Soar in Dubious Twitter Accounts Since Rakhine Attacks," Irrawaddy, Sept. 2, www.irrawaddy.com/news/soar-dubious-twitter-accounts-since-rakhine-attacks.html.

51 About Facebook website (2019), "Removing Myanmar Military Officials from Facebook," Nov. 7, https://about.fb.com/news/2018/08/removing-myanmar-officials/; Office of the UN High Commissioner for Human Rights (2018), Report of the Independent International Fact-Finding Mission on Myanmar, A/HRC/39/64 (OHCHR), https://www.securitycouncilreport.org/un-documents/document/ahrc3964.php.

52 Paul Mozur (2018), "A Genocide Incited on Facebook, With Posts From Myanmar's Military," New York Times, October 15, https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html.

53 Ronan Lee (2019), "Extreme Speech in Myanmar: The Role of State Media in the Rohingya Forced Migration Crisis," International Journal of Communication 13: 3202–3224, https://www.researchgate.net/publication/334762511_Extreme_Speech_in_Myanmar_The_Role_of_State_Media_in_the_Rohingya_Forced_Migration_Crisis.

54 Some extremist right-wing movements in the West effectively operate entirely online. See, for example, Alice Marwick and Rebecca Lewis (2017), Media Manipulation and Disinformation Online, (Data & Society), https://datasociety.net/output/media-manipulation-and-disinfo-online/.

55 Nuraniyah 2019 (note 23). This is also documented in some right-wing extremist circles: see Marwick and Lewis 2017 (note 54).

56 IPAC 2017a (note 3).

57 The Asia Foundation and Rappler 2018 (note 3).

58 Indonesia has also had cases in previous years of attempted recruitment through purely online channels, but, as in the Philippines, this is highly risky and a likely sign of a less-skilled, less-effective jihadist effort that will have a short life span before identification and crackdown by the authorities. Institute for Policy Analysis of Conflict (2015), Online Activism and Social Media Usage Among Indonesian Extremists, IPAC Report No. 24 (Jakarta: IPAC), http://www.understandingconflict.org/en/conflict/read/46/Online-Activism-and-Social-Media-Usage-Among-Indonesian-Extremists.

59 Institute for Policy Analysis of Conflict (2017b), The Radicalisation of Indonesian Women Workers in Hong Kong, IPAC Report No. 39 (Jakarta: IPAC), http://www.understandingconflict.org/en/conflict/read/62/The-Radicalisation-of-Indonesian-Women-Workers-in-Hong-Kong.

60 Brad Ridout et al. (2019), Mobile Myanmar: The Impact of Social Media on Young People in Conflict-Affected Regions of Myanmar (Yangon, Myanmar: Save the Children Myanmar and The University of Sydney), https://resourcecentre.savethechildren.net/library/mobile-myanmar-impact-social-media-young-people-conflict-affected-regions-myanmar.

61 See, for example, Audrey Alexander (2019), Perspectives on the Future of Women, Gender, and Violent Extremism, Occasional Paper Series (George Washington University), https://extremism.gwu.edu/sites/g/files/zaxdzs2191/f/Perspectives%20on%20the%20Future%20of%20Women%2C%20Gender%20and%20Violent%20Extremism.pdf.

62 Jytte Klausen (2015), "Tweeting the Jihad: Social Media Networks of Western Foreign Fighters in Syria and Iraq," Studies in Conflict and Terrorism 38 (1): 1–22, https://doi.org/10.1080/1057610X.2014.974948.

63 Pedro Manrique et al., (2016), "Women's connectivity in extreme networks," Science Advances 2 (6), https://advances.sciencemag.org/content/2/6/e1501742.full.

64 IPAC 2015 (note 58).

65 IPAC 2017b (note 59).

66 ICG 2017 (note 26).

67 ICG 2017 (note 26).

68 Schissler et al. 2015 (note 26).

69 IPAC 2015 (note 58). In a 2018 study, IPAC notes several specific functions for which extremists utilize social media platforms, including "propaganda, recruitment, fundraising, group reinforcement, dissemination of instructional material (including how to make bombs), and occasionally attack planning." Institute for Policy Analysis of Conflict (2018), Indonesia and the Tech Giants vs. ISIS Supporters: Combating Violent Extremism Online, IPAC Report No. 48 (Jakarta: IPAC), http://www.understandingconflict.org/en/conflict/read/72/Indonesia-and-the-Tech-Giants-vs-ISIS-Supporters-Combating-Violent-Extremism-Online.

70 "Rohingya Refugee Crisis," UNOCHA news release, accessed Dec. 10, 2019, https://www.unocha.org/rohingya-refugee-crisis.

71 International Crisis Group (2016), Myanmar: A New Muslim Insurgency in Rakhine State, ICG Report No. 283 (Brussels: ICG), https://www.crisisgroup.org/asia/south-east-asia/myanmar/283-myanmar-new-muslim-insurgency-rakhine-state.

72 IGG 2016 (note 71). ARSA was not, however, an international jihadist group, motivated instead by Rohingya conditions; its direct invocation of jihadist terms and rationale beyond a general Islamic identity was limited.

73 ICG 2016 (note 71). Although northern Rakhine has a history of militant movements built on Muslim Rohingya identity, ARSA personnel and identity appear entirely separate from these (indeed, their first video statement was to counter claims of responsibility for attacks by the remnants of the RSO, or Rohingya Solidarity Organization, established in the 1980s).

74 For platforms such as WhatsApp, where group members are generally added manually by administrators, this exclusivity is locked in. For others, such as Facebook, with both open and closed pages and news feeds that combine posts from friends with other material, the process superficially appears more organic, but restrictive selection still occurs.

75 Schissler et al. 2015 (note 26).

76 For more detailed exploration, see Gerard McCarthy (2018), "Cyber-Spaces," in Simpson et al. (eds), Routledge Handbook of Contemporary Myanmar (London and New York: Routledge); and Ridout et al. 2019 (note 60).

77 McCarthy 2018 (note 76).

78 Peter Cihon and Helani Galpaya (2017), Navigating the Walled Garden: Free and Subsidized Data Use in Myanmar (Colombo, Sri Lanka: LIRNEasia), https://lirneasia.net/wp-content/uploads/2017/03/NavigatingTheWalledGarden_CihonGalpaya_2017.pdf.

79 The term Burmese is used in their report to refer to the most widely used language of Myanmar (also known as Burma).

80 Bridget DiCerto (2014), "In Newly Liberated Myanmar, Hatred Spreads on Facebook," The World from PRX, Aug. 8, www.pri.org/stories/2014-08-08/newly-liberated-myanmar-hatred-spreads-facebook.

81 OHCHR 2018 (note 51).

82 Alex Warofka (2019), "An Independent Assessment of the Human Rights Impact of Facebook in Myanmar," About Facebook website, Nov. 5, http://about.fb.com/news/2018/11/myanmar-hria/.

83 See, for instance, Lovett 2017 (note 50).

84 The Asia Foundation and Rappler 2018 (note 3).

85 IPAC 2018 (note 69).

86 Other languages, like Rohingya, have no standard script and low levels of literacy among native speakers, resulting in some of the same effects.

87 Adriano and Parks 2013 (note 35).

88 The Asia Foundation and Rappler 2018 (note 3).

89 The Asia Foundation and Rappler 2018 (note 3). This does not mean there were no Filipino IS Telegram groups; there were at least 50 at the peak in 2016–17. Although some of them shared material that had been translated into Mindanao languages, however, the lack of Telegram use by the general public meant they could not connect to general local audiences.

90 Although now called IS or Daesh in most global coverage, this report—like many of the studies it draws from—uses ISIS, as it is the name more commonly used in Southeast Asia.

91 For a summary of the coalition, see, for example, IPAC 2016 (note 3).

92 The Asia Foundation and Rappler 2018 (note 3). This study took place a year after the battle. Pages therefore may have been deleted or taken down, although none of the literature consulted mentioned any such pages.

93 IPAC 2015 (note 58).

94 Indonesia's larger Islamic population also supports an extensive Islamic publishing and media industry, with skills flowing into extremist social media production. Nuraniyah 2019 (note 23).

95 47 U.S. Code § 230. Protection for private blocking and screening of offensive material, https://www.law.cornell.edu/uscode/text/47/230.

96 United States Court of Appeals for the Second Circuit, Force v. Facebook, Inc. No. 18-397, July 31, 2019, https://law.justia.com/cases/federal/appellate-courts/ca2/18-397/18-397-2019-07-31.html.

97 For a global account of intermediary liability laws and policies, see https://wilmap.law.stanford.edu/.

98 Timothy McLaughlin (2018), "How Facebook's Rise Fueled Chaos and Confusion in Myanmar," Wired, June, www.wired.com/story/how-facebooks-rise-fueled-chaos-and-confusion-in-myanmar/.

99 McLaughlin 2018 (note 98).

100 Sara Su (2019), "Update on Myanmar," About Facebook website, Aug. 15, www.about.fb.com/news/2018/08/update-on-myanmar/.

101 Samidh Chakrabarti and Rosa Birch (2020), "Understanding Social Media and Conflict," About Facebook website, May 21, www.about.fb.com/news/2019/06/social-media-and-conflict/.

102 Elise Thomas (2019), "Facebook Keeps Failing in Myanmar," Foreign Policy, June 21, www.foreignpolicy.com/2019/06/21/facebook-keeps-failing-in-myanmar-zuckerberg-arakan-army-rakhine/.

103 Facebook website (2020), "Facebook Transparency Report: Community Standards," May, http://transparency.facebook.com/community-standards-enforcement.

104 Louise Matsakis (2018), "QAnon Is Trying to Trick Facebook's Meme-Reading AI," Wired, Sept. 18, www.wired.com/story/qanon-conspiracy-facebook-meme-ai/.

105 No data was immediately available on how this affected ISIS supporters in Southeast Asia, but there is every reason to believe that it was similar to other regions. Rita Katz (2019), "ISIS Is Now Harder to Track Online—but That's Good News," Wired, December 16, https://www.wired.com/story/opinion-isis-is-now-harder-to-track-onlinebut-thats-good-news/.

106 Davey Alba (2019), "Ahead of 2020, Facebook Falls Short on Plan to Share Data on Disinformation," New York Times, Sept. 29, www.nytimes.com/2019/09/29/technology/facebook-disinformation.html.

107 Brent Harris (2020), "Preparing the Way Forward for Facebook's Oversight Board," About Facebook website, Jan. 28, www.about.fb.com/news/2020/01/facebooks-oversight-board/.

108 Facebook has said that in the future the board will also be able to review decisions not to remove reported content, but no timeline for this has been established. Review of the platform's ranking and recommendation algorithms does not appear on the board's roadmap. Harris 2020 (note 107).

109 David L. Hudson (2017), "Counterspeech Doctrine," in The First Amendment Encyclopedia (Middle Tennessee State University), www.mtsu.edu/first-amendment/article/940/counterspeech-doctrine.

110 John Shinal (2018), "Facebook, Google Tell Congress They're Fighting Extremist Content with Counterpropaganda," CNBC, Jan. 17, www.cnbc.com/2018/01/17/facebook-google-tell-congress-how-theyre-fighting-extremist-content.html.

111 Spandana Singh (2019), Everything in Moderation, (New America), www.newamerica.org/oti/reports/everything-moderation-analysis-how-internet-platforms-are-using-artificial-intelligence-moderate-user-generated-content/.

112 Mark Zuckerberg (2018), "A Blueprint for Content Governance and Enforcement," Nov. 15, https://www.facebook.com/notes/mark-zuckerberg/a-blueprint-for-content-governance-and-enforcement/10156443129621634/.

113 Schissler et al. 2015 (note 26).

114 Matthew J. Walton, Matt Schissler, and Phyu Phyu Thi (2017), Failed riots: Successful conflict prevention in four Myanmar cities, (Myanmar Media and Society Project), https://www.sant.ox.ac.uk/sites/default/files/m.mas_working_paper_1.2_-_failed_riots_-_january_2017.pdf.

115 Thant Sin Oo et al. (2019), "Exploring Digital & Mobile Cultures in Myanmar," (Phandeeyar), https://www.digitalculturesmm.com/index.html#method; Ridout et al. 2019 (note 60).

116 Schlisser (2015) also explores this in some detail at an earlier stage of Myanmar's rapidly expanding connectivity, noting that some online behaviors relate directly to the ways in which information was treated during the height of the military dictatorship, such as how rumors were relied upon and interpreted when other forms of information were unavailable. Schissler et al. 2015 (note 26).

117 Clément Silverman (2018), "Can positive messaging on social media promote peacebuilding in Myanmar?" (master's thesis, Malmo University), http://muep.mau.se/handle/2043/25017.

118 Ridout et al. 2019 (note 60).

119 Taylor O'Connor (2018), Stakeholder mapping of countering hate speech in Myanmar – External Report" (Washington, D.C., and Brussels: Search for Common Ground), https://www.sfcg.org/wp-content/uploads/2018/01/SFCG-Stakeholder-Mapping-Report-external-20Nov2017-FINAL-for-printing.pdf. This observation distinguishes Panzagar from many initiatives that have been observed to be poorly coordinated, although they were still felt to be effective. Ridout (2019) notes that "few participants had come across anti-hate-speech campaigns either online or offline; however, some noted that there were influential individuals who made efforts to counter and stop the spread of hate speech online." Ridout et al. 2019 (note 60). O'Connor also includes detail on common obstacles to conducting online anti-hate-speech initiatives in Myanmar, and a more detailed typology of interventions.

120 Silverman 2018 (note 117).

121 O'Connor 2018 (note 119).

122 Séraphin Alava, Divina Frau-Meigs, and Ghayda Hassan (2017), Youth and Violent Extremism on Social Media: Mapping the Research (UNESCO),

123 https://unesdoc.unesco.org/ark:/48223/pf0000260382.
For example, while Schissler (2015) mapped anti-Muslim narratives, researchers also noted that "we regularly encountered people who articulated feelings of contradiction between these narratives and their own experiences. They shared memories of times both recent and in the distant past in which they and friends, neighbors, and coworkers lived together. We think these kinds of memories are important and can be valuable components of peacebuilding and conflict prevention for at least two reasons. First, the existence of memories of peaceful interreligious coexistence can be the basis for prompting people to critically engage with currently circulating anti-Muslim narratives, including rumors, hate speech, or propaganda. Second, our research has uncovered instances where memories of peaceful coexistence or mutual assistance in the past seem to have been useful for people seeking to de-escalate tense situations they thought could erupt into wider violence or riots." Schissler et al. (2015, note 26).

124 Shiva K. Dhunguna (2017), A Rapid Assessment of the State of Conflict Dynamics and Information Management in Amarapura, Mandalay Division, and Lashio, Shan State, Myanmar (Yangon: Search for Common Ground), https://themimu.info/sites/themimu.info/files/documents/Report_Rapid_Assessment_of_Conflict_Dynamics_Information_Management_in_Mandalay_Shan.pdf.

125 IPAC 2018 (note 69).

126 Nuraniyah 2019 (note 23).

127 IPAC 2018 (note 69).

128 See, for example, Helen Pausacker (2019), "Baiq Nuril, the ITE Law and #MeToo Indonesian style," Indonesia at Melbourne (blog), July 18, https://indonesiaatmelbourne.unimelb.edu.au/baiq-nuril-the-ite-law-and-metoo-indonesian-style/; also Ika Ningtyas (2019), "Pressure on journalists," Development and Cooperation blog, Sept. 2, https://www.dandc.eu/en/article/defamation-clause-indonesias-internet-law-misused-criminalise-journalists.

129 Case study material from research undertaken by The Asia Foundation for a preliminary report on civil society engagement in preventing violent extremism in Southeast Asia, 2020. See also: Malaysian Islamic Youth Movement (ABIM) website, http://www.abim.org.my/about-profile.html.

130 Cynthia Dwork and Aaron Roth (2014), "The Algorithmic Foundations of Differential Privacy," Foundations and Trends in Theoretical Computer Science 9.3-4:211-407, https://doi.org/10.1561/0400000042.

131 Madeline Lamo (2018), "Regulating Bots on Social Media Is Easier Said Than Done," Slate, Aug. 9, https://slate.com/technology/2018/08/to-regulate-bots-we-have-to-define-them.html