© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

CONTENTS

Acknowledgments ix

1 Introduction 1

2 A Theory of Digital Resilience 20

3 Regulating Social Media Platforms 45

4 Adapting to Moderation through Migration 69

5 Safe Havens for Mobilizing and Recruiting 99

6 Reaching the Masses by Changing Messaging 122

7 Is Convergence in Moderation a Solution? 148

8 Conclusion: The Future of Online Harms 168

9 Appendix 181

Notes 199

Works Cited 235

Index 247

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

1

Introduction

ON FEBRUARY 3, 2015, the Islamic State (IS), a violent extremist organization based in the Middle East, posted a video on YouTube that showed the killing of a Jordanian pilot the group held hostage. The video depicted brutal acts of violence carried out against the hostage and called for the murder of anyone who did not agree with the group's mission.[1] Within a short amount of time, the video went viral. It was shared in tens of thousands of posts on Twitter by Islamic State sympathizers,[2] received millions of interactions on Facebook,[3] and was reposted and discussed in other online media.[4] The video was just one of many other videos showing brutalities that the organization posted on the internet. Thousands of other messages glorifying violence were shared by the group on various online platforms in an effort to attract support around the world.

Harmful content on social media, like the Islamic State's video described here, has become an urgent challenge for societies in the digital age. The stories seem to never end. From hate-filled campaigns that can lead to violence and even genocide[5] to extremism and polarization fueled by misinformation and manipulated media generated by artificial intelligence,[6] to problematic content that targets children or can inspire self-harm,[7] safety problems on online platforms have become a pressing concern.

As disturbing episodes linked to harmful content multiplied around the world, public calls for immediate action from social media companies started to intensify. Some threatened to boycott social media, calling on technology companies to do more to address harassment and abuse on their sites.[8] Others, such as large advertisers, withdrew their paid ads to protest platforms' insufficient handling of misinformation and hate speech.[9] And those who became victims of violence that was believed to be inspired by content posted online demanded legal action against companies that did not sufficiently combat incitement on their platforms.[10]

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

The most intense pressure came from governments. Between 2018 and 2022, forty-seven national governments enacted legislation or were in the process of passing legislation to regulate harmful content on social media. The regulations sought to address a range of social ills for which digital media platforms were believed to be responsible, and they reflected governments' view that leaving content moderation solely in the hands of private companies was insufficient to combat harmful content online.

For example, after the livestreaming on Facebook of the terrorist attack in Christchurch in March 2019, the prime minister of New Zealand said that "we cannot simply sit back and accept that [social media] platforms just exist and that what is said on them is not the responsibility of the place where they are published."[11] The prime minister of Australia agreed, adding that "big social media companies have a responsibility to take every possible action to ensure their technology products are not exploited by murderous terrorists."[12] Indeed, when crafting the Digital Services Act—the European Union's ambitious legislation to combat harmful content online—one regulator warned that "the time of big online platforms behaving like they are 'too big to care' is coming to an end."[13]

To put pressure on social media companies to do more, governments started requiring a range of actions. They asked companies to publish transparency reports detailing their content moderation actions and to set up measures to protect against misuse. They also required platforms to engage in risk assessments and audits to ensure that they were not exploited to promote offline harm. Governments further dramatically increased their requests for removal of specific pieces of content that they considered harmful, in hopes that tighter policing of speech would lead to a safer internet.[14]

But an interesting aspect of the public's attention and the regulatory pressure placed on technology companies was the almost exclusive focus on large social media platforms. Technology companies like Meta, Twitter, and YouTube faced strong public pressure, while smaller sites such as Telegram, Gab, and Rumble were subject to much less oversight. There are various reasons for this differential treatment of platforms, including efforts to lower the barrier to small social media companies to enter the digital media market and the view that smaller platforms are inconsequential.[15] Regardless of the reason, the result was a diverse, uneven moderation landscape in which social media platforms that were considered "big" or impactful moderated content at increasingly higher rates, while smaller platforms did not.

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

For example, in the past several years, Meta, which faced the highest public pressure among all technology companies, has developed over twenty categories for harmful content and removed over 64 billion posts promoting misinformation, hate, and extremism from its services.[16] TikTok, which also started experiencing pressure from regulators and the public, has increased its moderation substantially since 2020, taking down over 1.2 billion videos that violated its new policies.[17] But other technology companies that have not been at the center of the public's attention, such as Gab, Rumble, and Telegram, had much less robust content moderation standards and did not remove harmful content to the same extent.

How effective is content moderation at combating online harms in such diverse environments? We currently have little knowledge on the effects of content moderation when different technology companies adopt different standards to restrict harmful content on their sites. Most scholarly work on the regulation of social media either focuses on the legal aspects without examining the consequences of regulatory provisions or examines platforms' content moderation from a "within-platform perspective"—that is, by evaluating the effectiveness of moderation only among platforms that engage in it. As a result, we have a limited understanding of the effects of content moderation in the broader online ecosystem. What happens when one social media platform adheres to regulations and moderates harmful content and another does not? How do producers of harmful content adapt to moderation across platforms with different moderation policies? And what happens when social media companies collaborate to align their moderation standards?

I answer these questions by focusing on one of the most central areas in social media regulation: online extremism. Nowhere have public pressure and regulatory efforts been more prominent than the moderation of extremist and terrorist activity on digital media platforms. "Dangerous organizations"—militant or hate-based groups, extremist organizations, and other violent movements—have become one of the main targets of online regulation and content moderation. But despite mass content takedowns, account suspensions, and other sanctions that these organizations have been subject to, they continue to flourish online, advancing their cause, recruiting supporters, and inspiring violence. The new policies targeting harmful content on social media have thus not stopped extremist organizations from operating online. Why?

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

Moderating Harmful Content in a Diverse Online Environment

In this book, I provide a deep dive into content moderation on social media platforms to study how extremist organizations that produce harmful content online react to moderation. I define content moderation as the "organized practice of screening user-generated content (UGC) posted to Internet sites, social media, and other online outlets, in order to determine the appropriateness of the content for a given site, locality, or jurisdiction."[18] I argue that divergence between platforms' content policies allows these actors to become resistant to online regulation. Focusing on the strategic interaction between technology companies and national governments, I explain how different moderation standards emerge in the online information ecosystem. I then offer a theory that explains how different moderation standards across platforms create virtual safe havens in which extremist actors can organize, launch campaigns, and mobilize supporters. Drawing on data on the online activity of over a hundred militant and hate organizations, archives of banned terrorist propaganda, and platform moderation policies, I explain how digital resilience is shaped by the degree of variation in the way technology platforms police speech online.

Understanding how extremist actors adapt to moderation sheds light on important challenges at the frontier of mitigating online harms. Divergent standards in content moderation are a feature of our increasingly decentralized online information ecosystem, yet their effects on the ability to moderate harms are rarely considered in debates over social media regulation. Policymakers often rush to suppress or take down offensive content online, while failing to consider the consequences of these approaches in the broader digital environment. By explaining the ways in which variation in content moderation across social media platforms can be exploited by militant and hate organizations, the book provides an important account for why extremism continues to be a problem for our digitally connected societies.

Why Does Extremism Thrive Online?

A key question driving governments seeking to regulate internet platforms is determining how extremist actors are able to exploit social media for malicious purposes. The oft-cited answer is that social media companies are "not doing enough" to moderate harmful content—which naturally leads to

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

policy solutions requiring more moderation. In this book, I show that this answer is incomplete, because it ignores the broader online ecosystem in which social media platforms are embedded. By taking a more comprehensive, cross-platform perspective to answer this question, I illustrate the important role played by the *structure* of the online information environment and the way platforms relate to each other in shaping extremist organizations' online campaigns.

Specifically, I show that militant and hate organizations' online success centers on their ability to operate across many platforms in parallel—a phenomenon not well captured by current legislation. Most regulations of social media platforms operate under the assumption that inducing technology companies to take stronger action against extremist organizations will decrease their ability to exploit the internet. As a result, the main metric of success employed by governments and the public is a decline in harmful content on regulated platforms. What the new regulations tend to overlook is that these actors operate in tandem in multiple online spaces, many of which are unregulated.

My theory explains how dangerous organizations build resilience to content moderation by focusing in particular on a multi-platform environment. The context of my theory is an information ecosystem characterized by variation in the levels of platform moderation in which some online spaces have tight moderation of harmful content, while others do not. I show that much of the complexity in extremist group behavior on social media can be captured by two important dimensions in platform characteristics: (1) the level of moderation—how lenient or restrictive platforms are in their content moderation practices; and (2) impact—the size of the audience on the platform. This approach yields systematic and falsifiable predictions as to the mechanisms that facilitate dangerous organizations' resilience on social media.

The mechanism that most people point to when considering extremist groups' adaptation to content moderation is *migration* to alternative platforms. Faced with bans on moderated platforms, extremist actors can shift their online presence to social media spaces that have more lenient content policies. There are many examples of migration of this sort, including the Islamic State's move to Telegram after Twitter and Facebook banned the group from their platforms and the relocation of far-right groups to "alternative platforms" after experiencing crackdowns on mainstream social media sites. The ability to migrate to other platforms allows extremist organizations to maintain their presence online, but it comes with a cost. Since platforms that moderate less tend to have smaller user bases, extremist actors who migrate to less-moderated

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

online spaces risk becoming "irrelevant" by losing the audience they once had on larger platforms.

To remain influential in a regulated online context, extremist organizations have an incentive to continue reaching out to broader audiences, even on moderated platforms. I show that they do that by shifting *messaging* strategies based on platform moderation rules. When there is variation in the policies that determine what is allowed (and not allowed) on different platforms, extremist actors can modify the way in which they disseminate their messages across platforms. This tactic is particularly useful as content moderation becomes more automated, since it is often easier to "trick" artificial intelligence systems with borderline content that is not clearly violating platform policies. Thus, for example, these actors advance their message by sharing 'mild' content that does not trigger moderation algorithms on regulated platforms, while at the same time disseminating more extreme material in less-moderated spaces. Inconsistency in content moderation policies across platforms allows dangerous organizations to maintain a significant level of audience reach despite increased regulation.

The third mechanism for adapting to content moderation is *mobilization*—the ability to draw audiences to support the group's cause. When dangerous organizations maintain a presence on several platforms in parallel, they are not only able to evade moderation and increase the probability that their content will flow to target audiences, but they can also engage potential recruits in a more effective way. It turns out that experiencing moderation on social media—for example, by having one's posts deleted or account suspended—can generate strong reactions among affected users that can propel engagement with extremist content. Followers of the Islamic State who experienced bans on Twitter and Facebook subsequently became much more motivated to seek out the group's online networks on less-moderated platforms.[19] Similarly, sympathizers of the far-right conspiracy theory QAnon, which was one of the narratives inspiring the January 2021 attack on the US Capitol, increased their engagement with QAnon content after their accounts were deleted from mainstream platforms.[20] Targeting users who experience moderation can be very beneficial for extremist organizations, as these audiences might be more susceptible to their narratives.

I offer rich empirical evidence from various sources, including data on extremist groups' online networks, archives of banned terrorist propaganda, and data on technology companies' enforcement actions, to show how cross-platform migration, messaging, and mobilization allow militant and hate

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

organizations to remain resilient to online regulation. I show that even though content moderation can succeed in reducing extremist presence on specific, highly moderated platforms, it often fails to address online harms in the broader internet ecosystem. I present data on the online activity of over a hundred groups that promote extremist ideologies on various social media platforms, and I supplement these data with deep quantitative and qualitative case studies.

It is often difficult to study the consequences of content moderation because information on such interventions is usually not available to the public. In many cases, banned content cannot even be found in the archives of social media platforms because of data archiving laws and other restrictions. For this reason, I rely on several forms of external data collection that allow me to track instances of platform moderation and extremist activity in real time. I use these data sources to answer the following questions: Where do extremist actors migrate after experiencing bans on regulated social media platforms? What content do they post in different online spaces, and does content dissemination vary between moderating and nonmoderating platforms? Do users who experience bans increase their engagement with extremist content on unregulated social media sites? And how do extremist actors target audiences who are aggrieved by moderation? The evidence that I present demonstrates that even though content moderation can be effective in combating harmful content on specific platforms, its ability to prevent harm in the broader digital environment is more limited.

My findings help explain why online extremism—and harmful content more generally—continues to persist on social media. The political process leading governments to exert greater control over online platforms and the public pressure on technology companies to invest in content moderation are not uniform across social media platforms. This inconsistency creates opportunities for actors who produce harmful content to innovate in the online space, which can strengthen, rather than weaken, their online campaigns.

The evidence presented in this book demonstrates that we need to better understand the ways in which content moderation shapes dangerous actors' behavior in the broader digital ecosystem. Although there is a growing body of research on the movement of messages, ideas, and different types of content between social media platforms, we know much less about the nature of cross-platform spillovers in the context of content moderation—and in particular the role of these spillovers in extremist organizations' strategic attempts to overcome regulation.[21] This book shows that the movement of harmful con-

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

tent between platforms is often the outcome of more calculated efforts by these actors to evade moderation while maintaining their impact on moderated platforms. With an ever-expanding set of policies dictating what is not allowed on social media, understanding how extremist actors adapt to new rules can ensure that content moderation is more effective at achieving these important policy goals.

Defining Extremism

Before we continue, it is important to define "extremism," as the concept can have different meanings in different contexts. Scholars have long debated the meaning of extremism and have devoted entire articles or even books to discussion of the concept. Here I describe three broad definitions on which I draw in the book. The first defines "extremism" as a set of ideas, beliefs, or preferences that diverge from the views of the majority of the population.[22] If political preferences can be placed on a spectrum, then extremism would reflect the preferences on the far end. For example, organizations that promote fundamentalist religious beliefs that are not endorsed by the majority of a religious population would be considered extremist under this definition.[23] Similarly, groups that promote the transformation of a political system in a way that disrupts the popular status quo would be defined as extremist.[24]

The second definition focuses on intergroup relations. According to this definition, extremism is viewed as the belief that survival depends on taking a hostile action against others.[25] Organizations that promote hostility, such as verbal attacks, hate speech, or harassment, or even physical discrimination and violence against other groups are defined as extremist. Although extremism can emerge in many types of social contexts, the most common cases relate to groups that define themselves on the basis of religious, nationalist, racial/ethnic, or class-based identities.[26]

The third definition views extremism as synonymous with violence. Groups or individuals that promote physical harm against people or property, such as indiscriminate violence or terrorism, are defined as extremist. The large literature on preventing or countering violent extremism (PCVE) often views extremism as closely tied to violence and physical harm. Thus, "violent extremism" and "extremism" under this definition are seen as the same thing.[27]

In this book I draw on all three definitions when studying activity on social media. I examine groups and movements that promote fringe ideas that are

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

not popular among the majority in their society but that do not directly endorse violence against civilians. But I also study organizations that openly use indiscriminate violence to achieve their goals. By taking this approach, I put together in one bucket many different types of organizations that are often not studied at the same time. The reason for adopting a broad definition of extremism is that these different organizations, despite espousing different ideologies and tactics, face similar constraints when using social media. The theoretical framework in my book applies to many different organizations that fall under various definitions of extremism.

In addition, variants of these definitions are used by social media companies at various points in time. As I show later in the book, different definitions of extremism even become part of the story when they result in diverging policies to address the activity of dangerous organizations across platforms.

Defining Harmful Content

In a similar manner, the book focuses on policy efforts to combat harmful content on social media. But what is "harmful content"? Who decides what information is "good" and what is "bad"? This is a controversial question that results in different answers, depending on who we ask. For example, in polarized societies—where there is great animosity between two opposing sides of the political aisle—each camp may consider the other's ideology, worldview, or policy positions harmful. In a recent poll studying polarization in the United States, over 55 percent of the respondents believed that the other side's policy positions were subverting American democracy, and about 20 percent supported freezing the social media accounts of journalists who identified with the opposing party.[28] In another study examining public attitudes toward content moderation, researchers found a large difference in the views of Republicans and Democrats on whether news articles promoting misinformation should be removed from social media.[29]

The belief that online content is harmful if it does not support one's political goals is also found in other contexts. For example, in civil war and intrastate conflict, online content produced by nonstate actors is often considered terrorist propaganda by the governments fighting them, but some parts of the population that these organizations seek to represent view this content as resistance to a repressive regime. Disagreement over what constitutes harmful content is also found in debates over vaccine safety, where pro- and anti-vax activists see each other's online campaigns as harmful content.[30] Similar

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

dynamics exist in disagreements over political advertisements in the context of democratic elections, where opposing sides consider the others' campaigns misinformation.[31]

Here I avoid making normative judgments about what is "good" or "bad" content by focusing on information that national governments and technology companies have themselves deemed to be harmful in regulations and content moderation rules. My theoretical framework is thus agnostic on the exact definition of harmful content, allowing different stakeholders to define it according to their political views. Approaching this question in a liberal way allows me to directly examine a core aspect of my theory: heterogeneity in the ways that technology platforms define online harms. It also enables examination of policy evolution where the definition of harmful content changes over time.

For example, the QAnon conspiracy theory was not considered harmful by technology companies when it started spreading from fringe internet forums into mainstream social media. Platforms such as Facebook, Twitter, and YouTube did not include conspiracy theories in their definition of harmful content until QAnon communities grew significantly on their sites (see chapter 6). But at the same time that these platforms defined the QAnon ideology as harmful, other platforms, such as Gab, took a different position, maintaining that there was "nothing wrong" with the QAnon philosophy.[32] Embracing the varying definitions of harmful content allows me to study how QAnon supporters advanced the conspiracy's message on different platforms, thus testing important empirical implications of my theoretical framework.

Sometimes, however, there is broad consensus on certain types of harmful content. For example, many agree that content that promotes violence and terrorism is harmful.[33] In this book, I focus on extremist organizations—actors that, in many cases, promote violence to achieve political goals. Thus, the harmful content that I analyze often consists of messages shared on social media (in textual, visual, or audio format) that can cause injury, suffering, distress, or trauma outside of the platforms on which they are disseminated. This includes content that can inspire or instruct people to engage in violence, messages that promote hate and can lead to harassment of people, as well as information that can inspire individuals to harm themselves, such as content promoting suicide.

But even when there is agreement about what is harmful content, there can be disagreements about whether it should be moderated. For example, the leadership of Telegram has long agreed that content promoting violence and terrorism is harmful.[34] Despite this view, it did not put in place policies that

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

officially prohibit such content, citing its belief that user privacy is "more important than our fear of bad things happening, like terrorism."[35] Whether differences in content moderation across platforms stem from disagreement over what is harmful content, from different views about how it should be moderated, or from varying regulatory pressures, the outcome is the same: an uneven online information ecosystem in which some platforms have tight moderation rules and others do not. The discussion in this book thus applies equally to all types of scenarios: by taking variation in content moderation policies as given, it examines how heterogeneity in moderation standards shapes extremist activity online.

Contributions to Existing Literature

This book contributes to several strands of scholarly literature. First, it advances knowledge on the digital battleground where dangerous organizations operate. Over the past decade, the rise of internet platforms has inspired a wave of new scholarship examining how extremist actors use social media. This research has produced valuable insights into the ways in which militant and hate organizations produce propaganda, attract recruits, plan violent attacks, and—importantly—adapt to content moderation.[36] Although this work identifies important patterns in this book's outcomes of interest—for example, migration between platforms or the strategic modification of propaganda messages—it does a poor job of explaining the "independent variable," that is, the factors that give rise to these outcomes. In other words, existing research often treats the online information environment as having a "disorganized content moderation system." But as I show in this book, this system actually has predictable structural properties that are driven by the strategic incentives of technology companies and governments. By offering a framework to understand the properties of this environment and how it shapes the behavior of dangerous organizations, the book adds to knowledge on the causes of extremists' resilience on social media.

For example, it is well known that actors who are banned from one social media platform will try to migrate to another,[37] However, the literature has been underdeveloped on the incentives driving such migration and on where banned actors are most likely to end up. I draw on novel data on the online activity of a large number of militant and hate groups to show that, contrary to conventional wisdom, extremist actors do not migrate to the platforms that moderate the least. Rather, the choice is a trade-off: they are drawn to

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

platforms that reach a broad audience so long as the degree of moderation is tolerable. The book therefore offers new insights into the decisions that drive dangerous organizations' resilience on social media by systematically characterizing the online information environment in which these actors operate.

Relatedly, while we have research on how extremist mobilization takes place on social media, we do not have a good understanding of how content moderation policies drive mobilization *across* platforms.[38] Since the bulk of current research on extremist activity online focuses on a handful of platforms, the broader structure of the information environment and its influence on extremist activity is not systematically studied or theorized. By focusing on a wide range of social media platforms (including less-researched online spaces) and providing a framework that allows systematic characterization of their affordances, the book offers new insights into the consequences of content moderation—insights that allow us to make policy-relevant predictions about when interventions to combat harmful content online are most likely to succeed or fail.

Furthermore, the book sheds light on how changes in the regulatory landscape are likely to shape dangerous actors' online behavior in the future. Chapter 7 and the conclusion discuss policy responses to online extremism in light of the structural factors driving this information ecosystem. I provide novel evidence about what happens to extremist activity online when platforms converge in their policies to combat harmful content. Although the importance of collaboration between platforms to combat extremism is widely acknowledged, there is no systematic empirical research on this issue. The theoretical framework and empirical analysis offered in the book yield specific implications for regulatory safeguards that go well beyond our existing understanding of how extremist groups operate on social media.

Beyond online extremism, the book also adds to the growing body of work on censorship in the digital era. This stream of research—which often centers on digital repression in authoritarian regimes—points to different factors that allow citizens to become resistant to information suppression.[39] But censorship is not limited to authoritarian countries. Democratic governments are increasingly looking for ways to use similar tools to combat content that they consider harmful in their societies as well. This book explains how democratic governments grapple with the tension between censorship and free speech when faced with violence, hate, and extremism on social media platforms.

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

Finally, the book adds to work on resilience to online censorship—which tends to focus on information *consumption*—by providing a theoretical framework to understand how information *production* changes in the face of growing censorship. As such, the book pushes forward the frontier of knowledge on how actors who face bans on social media continue spreading their message online despite growing restrictions.

Implications for Mitigating Harms in a Digital Society

The theoretical argument and empirical findings that I present in this book shed light on the consequences of online regulation and content moderation and have implications for ongoing research on social media and politics, political violence, and the wider research agenda on online influence operations.

Content Moderation May Not Weaken Extremists

First, this book speaks to the strategies used by militant actors to exploit the internet to advance their cause. Many scholars have been puzzled by the ease with which groups like the Islamic State, Al-Qaeda, and other far-right organizations have used social media to recruit supporters and inspire violence around the world. Some argue that militant groups' creative online campaigns and sleek propaganda allow them to publicize their cause and successfully attract supporters on social media platforms.[40] Others maintain that extremist actors' capacity to effectively target audiences facilitates their success online.[41] Still others point to "offline" drivers of online extremism, highlighting militant and hate groups' ability to draw on grievances from the physical world to inspire audiences to engage with radical virtual communities.[42]

In this book I show that extremist groups have developed new methods to exploit the internet that go beyond propaganda dissemination, audience targeting, and offline grievances. The new information environment, which consists of various platforms that significantly differ in their content moderation standards, allows extremist actors to become resilient to online regulation. By migrating from platforms with strict moderation to those with more lenient policies, adapting messaging to platform rules, and mobilizing supporters on alternative platforms, militant and hate groups are able to continue operating successfully on social media, despite efforts to disrupt their campaigns.

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

The Broader Online Ecosystem Matters

This book also speaks to the growing body of research on social media and politics, which has made significant strides in understanding how rapid information dissemination in online networks shapes political behavior in the digital age.[43] The vast majority of this research tends to focus on large platforms, such as Twitter, Facebook, and YouTube.[44] But as I show in this book, the important dynamics on smaller platforms can be missed if online behavior is understood solely from the perspective of large social media platforms.

For example, when scholars study topics like online polarization, the spread of misinformation, and foreign and domestic influence operations, depending on whether they examine mainstream platforms, "alternative" platforms, or more private online sites, they may observe very different patterns. Researchers studying misinformation in India, for example, found that the vast majority of content promoting falsehoods was shared in small groups on encrypted messaging applications.[45] Scholars examining the spread of hate speech similarly found different patterns on alternative online networks, as opposed to mainstream platforms.[46] And studies on foreign influence operations identified similar activities on smaller online sites. For example, an investigation of Russia's influence operations targeting audiences in Europe and the United States found that networked accounts detected on Facebook were just the "tip of a much larger iceberg" of a campaign spanning many other, less-studied platforms.[47]

Understanding online behavior in the broader internet ecosystem is particularly important in the context of online extremism, as a growing portion of the activities of militant and hate groups take place on less-moderated sites. Indeed, many recent violent events have been linked to content shared on such platforms, including the terrorist attacks in France, Germany, and Belgium in the mid-2010s (facilitated by information shared on Telegram),[48] the 2018 Pittsburgh synagogue shooting (linked to Gab),[49] and the January 6, 2021, attack on the US Capitol (driven by misinformation related to the 2020 elections that spread on platforms such as MeWe, Gab, and Parler).[50] It therefore becomes increasingly important to expand our understanding of online political behavior by studying less-moderated platforms.

More generally, taking a cross-platform perspective can help explain the behavior of social media companies. For one, platforms may benefit from knowing where they are located in the broader online ecosystem; this information can help them better anticipate activities like sudden migration from other platforms. In addition, examining social media companies' content moderation

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

decisions in light of the broader online ecosystem can help explain why some platforms may choose to situate themselves in a place that captures the right mix of authenticity and impact—to attract extremists and grow their user base.

More Centralized Moderation Can Help, but It Has Costs

Beyond advancing knowledge on online behavior in the broader digital ecosystem, the book also speaks to the benefits and costs of cross-platform cooperation in the context of content moderation. In the face of growing urgency to combat online harms, many have called for greater cooperation between platforms, asking technology companies to work together to collectively prevent malicious content from spreading on the internet.[51] These efforts have propelled forward new initiatives, including New Zealand's Christchurch Call to Action, which over fifty countries have joined, and the Global Internet Forum to Counter Terrorism (GIFCT), which facilitates collaboration between online service providers in the removal of extremist content.

While industry collaboration in combating online harms is often cited as the primary solution to online extremism, I show that centralizing moderation across platforms also has costs. Drawing on a growing body of scholarship on online regulation and platform governance, I show that aligning moderation policies can indeed help technology companies create a "united front" against extremist exploitation. At the same time, however, it can give too much power over online speech to a small number of actors, potentially leading to abuse, as some have cautioned.

For example, centralizing content moderation across platforms can result in "collateral damage"—the mistaken removal of nonharmful content from many online sites in parallel. Since the line between violating and nonviolating content can be blurry at times, platforms sometimes make errors in their moderation decisions. When the moderation policies of technology companies converge, the impact of erroneous content removals can be even harsher, as those whose content is mistakenly deleted from one platform can experience bans on other platforms as well. Using examples of digital campaigns by civil rights organizations seeking to document human rights abuses around the world, I discuss centralized moderation's capacity to limit nonharmful actors' ability to share their messages online and the ethical questions about banning protected speech through such cross-platform collaboration that arise.

The book discusses other caveats to multi-stakeholder coordination in moderation, including "censorship creep"—a problem identified in the literature in

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

which moderation extends beyond its initial set of categories and centralization tools are misused by authoritarian leaders seeking to suppress dissent.[52] Thus, greater alignment in content moderation can facilitate the efficient removal of extremist material from social media platforms, but it comes with a cost to human rights and civil liberties, at least in some settings.

Relevancy to Other Online Campaigns

Finally, even though this book's focus is on militant and hate organizations, its findings are relevant for understanding other actors' adaptations to online regulation. Indeed, building resilience to content moderation is not unique to violent extremist groups. Even state actors seeking to influence domestic politics in their country or interfere in the political processes of other countries increasingly rely on multi-platform operations.

For example, a study by the Atlantic Council's Digital Forensic Research Lab found that a Russian intelligence campaign seeking to interfere in the domestic politics of the United States, Germany, and Ukraine (among other countries) spanned over thirty social media platforms and nine languages.[53] To evade moderation and increase impact, the operation employed "fake" accounts and impersonated user identities—a sophisticated messaging strategy that, as I show in the book, is often a part of more general efforts to overcome moderation.

Similarly, an influence operation by the Nicaraguan government in the run-up to the November 2021 elections employed hundreds of accounts on seven platforms to shape domestic audiences' vote choice. While some platforms, such as Facebook, Twitter, and YouTube, took action against this content, other platforms, like Telegram and TikTok, did not ban the campaign.[54] These examples illustrate the important role played by cross-platform variation in content moderation in the ability of malicious actors to spread their message online. By providing a framework for understanding these actors' adaptations to moderation in the broader internet ecosystem, the book illuminates often-overlooked consequences of online regulation.

Outline of the Book

The next chapter offers a theory that explains how policy divergence in content moderation leads to digital resilience. The theory focuses on the strategic interaction between national governments, technology companies, and dangerous organizations and explains the mechanisms that allow extremist groups to

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

continue advancing their goals online despite rising regulation. Focusing on democratic countries, I argue that governments develop policy interventions compelling content moderation by technology platforms—which is in some tension with liberal values of freedom of expression—when they perceive the risk of online harms crossing a certain threshold. Because this threshold is informed by activity taking place on large and high-impact social media platforms, government regulation tends to focus on this portion of the online information ecosystem. The result is variation in moderation levels across platforms: while larger, highly regulated platforms invest in moderation policies and enforcement, smaller, less-regulated platforms have a freer hand.

I argue that when social media platforms differ in their content moderation policies, militant and hate groups adapt through *migration*—moving from highly moderated platforms to less-moderated ones; *mobilization*—leveraging audience grievances from content bans to inspire support for their cause on less-regulated spaces; and *messaging*—adapting content to the moderation rules of different platforms. The key insight that emerges from my theoretical framework is that variation in content policies between platforms is central to extremist organizations' ability to build resistance to moderation.

In chapter 3, I provide an overview of recent efforts to regulate social media platforms to combat online harms. By tracking the evolution of social media regulation in various countries around the world, I show how governments' push to moderate content is closely related to activity that is perceived as "too harmful" because it is violent, threatening, or disruptive to democratic processes. I further show that regulations, or proposed regulations, tend to focus on "high-impact" platforms, and demonstrate how government pressures incentivize technology companies to invest in moderation. The chapter shows that when extremist material online is at issue, many democratic governments do not hesitate to mandate strong censorship that delves deep into the specifics of how private companies should moderate content.

Chapter 4 is the first of three chapters that provide evidence on the mechanisms that enable dangerous organizations to become resilient to moderation. In this chapter, I focus on cross-platform migration. Building on the theoretical framework in chapter 2, I show that two factors drive extremist groups' migration between platforms: the level of moderation and audience reach. Drawing on original data on the social media activity of over a hundred militant and hate organizations, I demonstrate that when choosing platforms, groups tend to gravitate to online sites that provide the best "mix" of moderation and reach.

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

The chapter also includes two case studies that trace the migration of militant and hate groups between platforms. First, I use data on the Islamic State's networks on Twitter and Telegram to show that the enforcement of moderation policies that targeted the group in 2016 and 2019 led its members to migrate to less-moderated platforms that served as safe havens for its online campaigns. Second, I describe the migration decisions of the Proud Boys in the wake of their deplatforming from mainstream social media, showing the important role of lenient moderation and audience size in the group's migration decisions. This chapter sheds light on current debates around deplatforming by not only illustrating how migration happens but also explaining where banned actors are most likely to move.

In chapter 5, I discuss dangerous organizations' use of less-moderated platforms for mobilization and recruitment, focusing on the audience from which extremist groups seek to mobilize supporters. I argue that even though less-moderated platforms have smaller audiences, they can be useful for mobilization when the content disseminated by militant and hate groups appeals to their users. I show that individuals' personal experiences with content moderation can make them susceptible to propaganda, especially when propaganda content effectively targets their grievances from moderation. Drawing on novel cross-platform data on social media users' reactions to deplatforming, I show that militant groups leverage frustration from moderation to attract support for their cause. In particular, I focus on the case of the Oath Keepers' mobilization on less-moderated platforms in the wake of the 2020 US elections to demonstrate how frustration from "Big Tech censorship" attracted users to the group's online campaign to "Stop the Steal" and "Defeat the Coup"—a campaign that flourished on less-moderated social media sites in the leadup to the attack on the US Capitol on January 6, 2021.

Although dangerous organizations commonly use less-moderated platforms as targets for migration and mobilization, these actors also often seek ways to reach audiences on larger, more moderated sites. In chapter 6, I explain how militant groups overcome moderation even on highly regulated platforms by changing how they communicate. First, I show that these actors often sacrifice the "authenticity" of their message by toning down violating content in order to gain greater impact. I demonstrate this adaptation process with data on the Taliban's networks on Twitter, which shows how the group strategically shifted its messaging to adapt to the platform's rules banning the glorification of violence. I further show that adaptation to moderation takes place across platforms by presenting a study of the QAnon movement's messaging shifts

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

on Twitter and Gab in response to changes in moderation rules. A second adaptation strategy involves "smuggling in" prohibited messages under the guise of content that appears innocuous. I argue that regulatory pressures to make platforms' content moderation policies transparent and accessible have the unintended consequence of allowing groups to more easily find ways to evade moderation. Drawing on examples from the Islamic State's evasion of moderation on Facebook, I describe how understanding moderation practices allows groups to build resilience, even on large regulated platforms.

In chapter 7, I discuss the role of interplatform cooperation in preventing dangerous organizations' exploitation of social media. Returning to the theoretical framework, I ask the following question: If divergence between platforms' content moderation policies increases extremist actors' resilience to moderation, does convergence weaken their ability to do so? I argue that alignment in moderation across platforms can limit migration and mobilization on smaller platforms and—to some extent—mitigate evasion on larger platforms as well. I examine trends in moderation alignment by drawing on original data that track changes in the content moderation rules of sixty social media platforms. I show that convergence in moderation across platforms is associated with lower levels of exploitation by extremist groups; in particular, convergence motivates a shift away from migration and mobilization on smaller platforms toward greater messaging adaptation on larger sites. The chapter further discusses the trade-offs of centralized moderation—in particular, its negative impact on nonharmful speech, its threats of over-censorship, and its ability to facilitate the creation of "content cartels."[55] As such, the chapter returns to a theme discussed in the beginning of the book: the role of moderation and censorship in democratic societies.

Chapter 8 concludes the book with a discussion of the broader implications for social media regulation in a multi-platform environment. Departing from the book's focus on extremist organizations, I consider the relevance of the patterns presented in the book to other settings, such as mis- or disinformation campaigns and state-sponsored influence operations. I outline directions for future research on the effects of platform moderation and discuss the potential for online regulation to transform our understanding of media and political behavior in the digital age.

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

INDEX

Italic pages refer to figures and tables

abuse: convergence and, 165; CSAM and, 173–76; digital resilience and, 20; future issues and, 173–76; messaging and, 138; migration and, 86, 91; mobilization and, 99, 104, 115; moderation and, 1, 15; regulation and, 54; research methodology and, 187
accountability, 45, 51–52
Action Days, 91, 93
advertisements: Digital Services Act and, *63*; hate speech and, 1; migration and, 85, 90, 95; Oath Keepers and, 119; politics and, 10, 104; pulling back ads, 1, 42; QAnon and, 136; targeting by, 85, 95; technology companies and, 22
Afghanistan: Biden and, 127; Dari language, 129–30, 145–46, *192–93*; Kabul, 133, 224n36; Mazar-i-Sharif, 126–27; Pashto language, 129–30, 145–46, *192–93*; Taliban and, 126–27
Africa, *76*, 84
African Americans, 172, 187
algorithms: Adam optimization, 188; automated, 6, 39, 144, 164; convergence, 228; document topics and, 106; Facebook, 144; harmful content and, 115–16; hate speech and, 115–16; moderation, 6, 39, 144, 152, 213n88, 228n23; regulation and, 52
Al Qaeda: Al-Shabaab and, 37; convergence and, 166; digital resilience and, 37; Facebook and, 31, 50; Mapping Militants Project and, 183; messaging and, 128; migration and, 31; recruitment and, 13; regulation and, 50; Twitter and, 31, 128
Al-Shabaab, 38, 78, 128, 166, 183
Amazon, 144, *145*, *152*, 214n10
American Defense League, 172
antifa, 95, 97
API (application programming interface), 186, 220nn20–21, 234n1
Apple, 85, 214n106
Arabic, 129–30, 144, 154, *192–93*, 228n21
Arab Spring, 176
Ardern, Jacinda, 28, 151
artificial intelligence (AI), 1, 39, 49, 174, 178–80
Asians, 187
Atlantic Council Digital Forensic Research Lab, 16, 127, 204n54, 205n16
Atomwaffen Division, 78
Australia, 2, 65, 84, 94, *152*, 181
Austrian Communication Platforms Act (KoPL-G), 57
authenticity: convergence and, 149, 157, 163; digital resilience and, 22, 24, 26, 36, 39–43; future issues of, 169–72, 175; impact and, 15, 18, 22–26, 29–33, 43, 70–75, 79, 82, 91, 93, 98, 117, 149, 163, 169; keywords for, *196*; maximizing, 79–82; messaging and, 18, 22, 24, 26, 29–31, 43, 117, 123, 147, 169; migration and, 70, 73, 75, 79, 82, 91–93, 98; mobilization and, 117; moderation and, 15, 18, 24, 26, 29–33, 36, 70, 73, 79, 82, 91, 98, 117, 123, 147, 149, 163, 169–72

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

authoritarianism: censorship and, 39, 219n14; digital repression and, 12, 16; liberation technologies and, 176–77; Nazi Germany and, 54; regulation and, 45, 67; terrorists and, 166
automation: algorithms, 6, 39, 144, 164; crawlers, 228n20; digital resilience and, 27–28, 39; Facebook detection, 144; migration and, 83–86; mobilization and, 108; moderation, 6, 27–28, 39, 57, 85–86, 125, 138, 144, 228n20; recommendations, 138; text analysis, 187; trackers, 129
Avia Law, 58–59, 65, 181

Bahrain, 176
bans: convergence and, 160, 164, 170; digital resilience and, 21, 33, 36, 40; Facebook, 5–6, 40, 96–97, 100, 144; messaging and, 123, 128, 133, 143–45; migration and, 69, 72, 73–74, 79, 82, 93, 96–97; mobilization and, 100–4, 111, 116; moderation and, 5–7, 13, 15, 17; research methodology and, 185; smuggling content and, 144; suspensions and, 6 (*see also* suspensions); Telegram, 5, 36, 82, 93, 97; Twitter, 5–6, 40, 82, 93, 96, 100, 111, 116, 128, 133, 143
Belgium, 14, *152*, 181
bias, 34, 52, 103, 107, 114
Bickert, Monika, 49–50
Biden, Joe: Afghanistan and, 127; Big Tech and, 51–52; legislation and, 51–52; Oath Keepers and, 99–100, 118–19; Taliban and, 127
Big Tech: Biden and, 51–52; censorship and, 18, 52, 111, 113, 166, *189–91*, 221n34; migration and, 111–15
Bing AI, 179
BitChute, 31
blogs, 49, 139
bots, *107*, 108, 179, 228n20, 234n46
Bradford, Anu, 53
Brennan, William, 207n7
Breton, Thierry, 61, 178
Britain First, 31
Buchanan, Pat, 94
bullying: categorization of, 71, *72*; convergence and, *157*; keywords and, *195*; migration and, 71, *72*, 80, *81*; NGOs on, *185*, 186

Cedar Revolution, 176
censorship: Big Tech, 18, 52, 111, 113, 166, *189–91*, 221n34; China and, 34, 103; convergence and, 165–66; creep of, 15–16, 165; deplatforming and, 111–15; digital resilience and, 34, 39, 43; freedom of expression and, 17, 21, 23, 53, 56, 59–61, 67, 85; future issues and, 171, 176; messaging and, 123; mobilization and, 99–103, 107–8, 111–14, 121, 219n14; moderation and, 12–19; regulation and, 46, 48, 52, 56
Center for Countering Digital Hate, 172
ChatGPT, 179
child sexual abuse material (CSAM), 173–76
China, 34, 103, 167, 174, 181
Christchurch Call to Action, 15, 28, 44, 151–53, 170
Citizens Protection (Against Online Harm) Rules, 67
Citron, Danielle, 165–66
Clubhouse, 154
CNN, 84–85
Code of Conduct, 59–60
Code of Practice on Disinformation, 172
collaboration: Code of Conduct, 59–60; convergence and, 150–67, 170; digital resilience and, 28; limiting extremists by, 161–64; migration and, 89, 91; moderation and, 3, 12, 15; regulation and, 59–60
collateral damage, 15, 164
Communications Act, 47
communism, 94, 114, 167, 174, 183
conservative values, 52, 99, 114, 118
conspiracy: convergence and, 157; digital resilience and, 31, 34, 40; messaging and, 136–39, 143; misinformation and, *196*; mobilization and, 107–8; moderation

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

and, 6, 10; QAnon and, 6, 10, 34, 40, 108, 136–39, 143
content adaptation: convergence and, 148, 150, 162, 164; CSAM and, 175; digital resilience and, 21–26, 29–30, 35–38, 41, 43; Facebook and, 5–6, 19, 125, 174, 229n30; Gab and, 19, 140; harmful content and, 3–4, 8, 19, 29, 36–37, 41, 43, 123, 148, 168, 175; hate speech and, 17, 24, 29–30, 36–38, 43, 148, 164, 168; messaging and, 123–45; militant groups and, 17–18, 24, 29–30, 36–38, 43, 123, 134, 148, 164, 168; moderation and, 3–8, 11, 16–19; QAnon and, 134–43; spamouflaging and, 174; Taliban and, 18, 123–25, 128, 131, 134, 143, 145; Telegram and, 5, 36, 229n30; threshold evasion and, 143–46; Twitter and, 5–6, 18–19, 123, 125, 131, 134, 140–45, 229n30; unaffiliated accounts and, 144–46; YouTube and, 125, 174
content masking, 144, *145*, 174–75
convergence: abuse and, 165; algorithms and, 228; alignment in thresholds and, 148–50; Al Qaeda and, *166*; Al-Shabaab and, *166*; authenticity and, 149, 157, 163; bans and, 160, 164, 170; bullying and, *157*; censorship and, 165–66; China and, 167; collaboration and, 150–67, 170; conspiracy and, 157; content adaptation and, 148, 150, 162, 164; costs of, 164–67; cross-platform cooperation and, 150–55; dangerous actors and, 153, 158; deplatforming and, 162; digital resilience and, 150; disinformation and, *157*; divergence and, 19, 148–50, 160; extremists and, 148–57, 161–67, *197*; Facebook and, 151–53, 160–61, 165–66, 228n20, 228n24, 229n30; Gab and, 160, *161*, 167, 228n24; Global Internet Forum to Counter Terrorism (GIFCT) and, 153–54, 158, 162, 165–67, 227n13; Google and, *152*; governmental calls for collaboration, 151–53; harassment and, *157*, 167; harmful content and, 148–58, 161–66, 196, 227n11, 228n23; hate speech and, *157*, 158, 166; Instagram and, 228n20, 228n24; Islamic State (IS) and, 154, *166*; keywords and, 156, *157*, 164; legal issues and, *166–67*; lenient content policies and, 148–50, 160–61, 166; manipulation and, *157*, 167; measuring, 156–68, 196, *197*; messaging and, 148–50, 155, 161–62; Meta and, *152*; Microsoft and, *152*, 153; militant groups and, 148–49, 154, 161–67; misinformation and, 157; moderation threshold and, 148–50, 155–67; Nazism and, *157*; Parler and, 160, *161*, 167; Pinterest and, 160, *161*, 228n24; policy alignment and, 160; propaganda and, 154, 162–67; racism and, 157; recruitment and, 150; Reddit and, 228n24; suspensions and, 162, *163*, 166; technology and, 150–55, 160, 164–67, 196, 227n13; Telegram and, 160, *161*, 229n30; terrorists and, 147, 151–54, 166–67; TikTok and, 160, *161*, 228n24; transparency and, 153, 227n13; trends in, 158–60; Twitter and, 152–53, 160–61, 228n24, 229n30; United States and, *152*, 166; video and, 151; violence and, 151, 153, 157, 165; YouTube and, 151–53, 160, *161*, 228n24
Conversations, 87
COVID-19 pandemic, 137–38
Cox, Christopher, 48–49
crime, 55, 58, *72*, *157*, 183, *194–95*

dangerous actors: convergence and, 153, 158; digital resilience and, 32, 44; extremists and, 6, 12, 17, 19, 21, 44, 49, 81, 153, 157, 166, 170, 178, 194; future issues, 178; hate speech and, 17, 24, 44, 71, 80–81, 96, 98, 158, 166, 170, 185–86, 194; migration and, 71, 80–81, 96, 98; militant groups and, 17–18, 24, 32, 44, 71, 80–81, 166, 170; moderation and, 7, 12; research methodology, *185*, *194*
Dari, 129–30, 145–46, *192–93*
dark web, 25
Dawn of Glad Tidings, 83
Death of the West (Buchanan), 94
“Defeat the Coup” campaign, 18, 100, 118

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

democracy: freedom of expression and, 17, 21, 23, 47, 53, 56–61, 67, 85; future issues and, 176–77; liberal values and, 45, 67–68; moderation and, 9; regulation and, 45, 52, 65, 67–68
Democrats, 9, 52
deplatforming: censorship and, 111–15; convergence and, 162; extremists and, 18, 100–3, 108–21; Great Deplatforming, 102, 105; hate speech after, 115–16; influencers and, 13–16, 102; measurement of, 108; messaging and, 131, *132*, 139; migration and, 69–70, 78, 82; mobilization and, 100–21; moderation and, 18; platform usage patterns and, 110–11; radicalization and, 180; regulation and, 52; research methodology and, 108–10, *191–92*
digital resilience: abuse and, 20; Al Qaeda and, 37; Al-Shabaab and, 38; authenticity and, 22, 24, 26, 36, 39–43; automation and, 27–28, 39; bans and, 21, 33, 36, 40; censorship and, 34, 39, 43; collaboration and, 28; conspiracy and, 31, 34, 40; content adaptation and, 21–26, 29–30, 35–38, 41, 43; convergence and, 150; dangerous actors and, 32, 44; divergence and, 16–17, 20–21, *26*, 30–41, 44; extremists and, 20–44, 46, 204n2, 204n6, 206n26; Facebook and, 20, 27–28, 31, 34–35, 39–40; freedom of expression and, 21, 23; future issues and, 169, 176; Gab and, 20, 31; harmful content and, 20–29, 35–44; hate speech and, 20, 31, 36; information flows and, 24–26; Islamic State (IS) and, 31, 34, 37–41, 204n6; keywords and, 39; legal issues and, 23, 28–29; legislation and, 23, 29, 46; lenient content policies and, 23–26, 43–44; messaging and, 21, 29, 33–38, 41–44; migration and, 69, 93; militant groups and, 20–25, 29–44; misinformation and, 22, 42; moderation threshold and, 21–40, 43; national governments and, 23–24; Parler and, 31; propaganda and, 24, 37–42, 204n6; QAnon and, 34, 40; recruitment and, 22, 33, 206n20; regulation and, 46, 62; strategic behavior in, 21–24; suspensions and, 31–34, 43; Taliban and, 40; technology and, 20–27, 35–42, 204n2; Telegram and, 31, 36; terrorists and, 27–28, 38–42; theory of, 20–44; TikTok and, 36; transparency and, 29, 38–39; Twitter and, 28, 31, 34, 37, 40; United States and, 28, 40; video and, 22, 39; violence and, 28, 39, 42, 204n6; YouTube and, 28, 31, 37, 40
Digital Markets Act, 61
Digital Services Act (DSA), 2, 57, 61, *63*, 66, 181–82
Discord, 184; convergence and, 228n24; CSAM and, 175; migration and, 75–78, 87–88, 215n4
disinformation: Code of Practice on Disinformation and, 172; convergence and, *157*; Ethiopia and, 67; Europe and, 172; future issues, 19, 172–74; influence operations and, 13–16, 19, 124, 173–74; keywords and, *196*; US Congress and, 53
"Disinformation Nation: Social Media's Role in Promoting Extremism and Misinformation" (Congressional hearing), 53
Disinformation Prevention and Suppression Proclamation, 67
divergence: content adaptation and, 35–37; convergence and, 19, 148–50, 160; digital resilience and, 16–17, 20–21, *26*, 30–41, 44; future issues and, 169, 173; messaging and, 35–41; migration and, 30–33, 69; mobilization and, 33–35; moderation and, 4, 8–9; threshold evasion and, 38–41
Dropbox, 91, *92*
Durov, Pavel, 89

Egypt, 73, 176
8chan, 135, 151
Encyclopedia of Political Communication, 115

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

English, 228n21; banned PDF books and, 90; in evading moderation, 146; Islamic State (IS) and, 90, 144; migration and, 77, 90; Taliban and, 129–30, *192–93*
Eshoo, Anna, 51
ethics, 15, 66, 109, 164–65, 182
Ethiopia, 67
Europe: Breton letter and, 178; Code of Conduct, 59–60; Code of Practice on Disinformation, 172; Committee on Industry, Research, Energy, 61; Digital Services Act (DSA), 2, 57, 61, *63*, 66, 181–82; global view on, 62–64; hands-on approach of, 53–62; messaging and, 136; migration and, 78, 84, 90, 94; "one-hour rule" and, 59–60; regulation and, 53–62; Security Union, 60; TERREG and, 60–61; Ukraine and, 174; X and, 172–73
Europol, 90–93
extremists: convergence and, 148–57, 161–67, *197*, 228n24; dangerous actors and, 6, 12, 17, 19, 21, 44, 49, 81, 153, 157, 166, 170, 178, 194; database on, 183–84; defining extremism, 8–9; deplatforming and, 18, 100–3, 108–21; digital resilience and, 20–44, 46, 204n2, 204n6, 206n26; future issues and, 168–80, 230n1; keywords for harmful content, *194–96*; messaging and, 123, 128, 135, 137, 142, 147; migration and, 5, 7, 11, 13, 17–21, 30, 35, 44, 69–71, 74–77, 81–87, 100, 104, 123, 148–50, 161, 170; mobilization and, 100–2, 104, 108–10, 115–21; moderation and, 1–19, 148–57, 161–67; nationalism and, 8, 75, *157*, *194*; online thriving of, 4–8; PCVE and, 8; platform usage patterns and, 110–11; radicalization, 13, 33, 49, 51, 84–85, 177, 180; recruitment and, 6, 11, 13, 18, 33, 85, 118; regulation and, 46, 49–54, 57, 62, 67–68; Southern Poverty Law Center database of, 183–84; Thompson letter and, 28; violence and, 1, 3, 8, 11, 14–17, 28, 42, 49–51, 57, 62, 81, 85–86, 108–9, 128, 137, 151–53, 157, 165, *194*. *See also* specific group
Facebook: Afghanistan and, 127; algorithms of, 144; Al Qaeda and, 31, 50; audience size of, 73; bans and, 5–6, 40, 96–97, 100, 144; China and, 174; Christchurch terrorist attack and, 2, 27–28, 60, 151–53; Code of Conduct, 59; Community Standards, 71; content adaptation and, 5–6, 19, 125, 174, 229n30; convergence and, 151–53, 160–61, 165–66, 228n20, 228n24, 229n30; CSAM and, 175; digital resilience and, 20, 27–28, 31, 34–35, 39–40; future issues and, 168, 174–77; Global Internet Forum to Counter Terrorism (GIFCT), 153, 164; hate speech and, 14, 31, 54, 57, 59, 96, 101–2; Islamic State (IS) and, 1, 5–6, 19, 31, 39, 83, 85, 87, 91, 122, 144–46, 174; lenient content policies, 169–74, 177; messaging and, 122–27, 133, 136–39, 143–46; migration and, 70–78, 83–98, 215n4; mobilization and, 99–104, 117–18, 121; moderation and, 1–2, 5–6, 10, 14, 16, 19; Paris terrorist attacks and, 85; policy alignment and, 160; propaganda on, 39–40, 50, 83, 85, 100, 122, 125, 144–45, 165, 174; Proud Boys and, 94–97; QAnon and, 6, 10, 34, 40, 136–39, 143; regulation and, 49–50, 53–61, 214n106; research methodology and, 184; suspensions and, 34, 99, 125, 133, 166, 222n1; Taliban and, 125–26; Thompson letter to, 28; Trump and, 221n1; US Capitol attack and, 6, 34, 222n1; video and, 1, 39, 50, 83, 95, 136, 151, 175; white supremacists and, 27; YouTube and, 168, 174–78
fake accounts, 16, *72*, 73, 104, *157*, *196*
FBI, 86, 137
Federal Communications Commission (FCC), 47, 207n7
Films, Videos and Publications Classification Amendment Bill, 65
Fishman, Brian, 50–51
flagging, *63*, 85, 174, 213n88
Florida, 52–53, 210n44
4chan, 135, 178

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

France: Avia Law and, 58–59, 65, 181; Central Office for the Fight Against Crime Related to Information and Communication Technology, 58; Constitutional Council, 59; Macron and, 28, 59, 151; terrorists and, 14, 28, 58–59, 85
France24 channel, 144, *145*
Freedom House, 64
freedom of expression: censorship and, 17, 21, 23, 53, 56–61, 85; digital resilience and, 21, 23; liberal values and, 17, 21, 23, 47, 53, 56–61, 67, 85; migration and, 67, 85; moderation and, 17; regulation and, 47, 53, 56–61, 67
Friendica, 73
future issues: abuse, 173–76; AI, 174, 178–80; authenticity, 169–72, 175; censorship, 171, 176; dangerous actors, 178; democracy, 176–77; digital resilience, 169, 176; disinformation, 19, 172–74; divergence, 169, 173; extremists, 168–80, 230n1; Facebook, 168, 174–77; Google, 179; harassment, 172; harmful content, 168–80, 230n2; Instagram, 174–77; Islamic State (IS), 172, 174, 180; legislation, 177; messaging, 169–70, 177; Meta, 174–75, 178; Microsoft, 179; militant groups, 168–71; misinformation, 172, 179; moderation threshold, 174; propaganda, 168–74; Proud Boys, 180; QAnon, 180; recruitment, 170; suspensions, 170, 172; Taliban, 180; technology, 168–80, 230n2; Telegram, 174–77; terrorists, 177; TikTok, 174–78; violence, 177–78; X, 171–74, 177–78

Gab: audience size of, 73; censorship and, 111–15; content adaptation and, 19, 140; convergence and, 160, *161*, 167, 228n24; data collection on, 186–87, *189–93*; digital resilience and, 20, 31; Mastodon and, 186, 220n20, 234n1; messaging and, 139–42, 226n64; migration and, 73–79, 215n4, 218n2, 220nn20–21, 220n24; mobilization and, 99–100, 105–21; moderation and, 2–3, 10, 14, 19; Oath Keepers and, 99–100, 110, 115–21, *189*, 222n54, 222n58, 222nn61–62; platform usage patterns and, 110–11; Proud Boys and, 96–98; QAnon and, 10, 108–42, *193*; research methodology and, 184, 186–87, *189*, *191*, *193*; suspensions and, 105, 109–12, 116, 119, *191*; 2020 elections and, 14, 105, 115, 118, 121
Gemini, 179
Germany, *152*; Criminal Code and, 56; hate speech and, 14, 53–58; Merkel and, 54; moderation and, 14, 16; Nazi, 53; Network Enforcement Act (NetzDG), 29, 53–58, 62, 65, 181; terrorists and, 14, 55
Global Internet Forum to Counter Terrorism (GIFCT): convergence and, 153–54, 158, 162, 165–67, 227n13; cost of, 15; Facebook and, 153, 164; facilitating cooperation and, 153–54; harmful content uploads and, 38; Microsoft and, 153; New Zealand and, 15, 166; Twitter and, 153
Google: convergence and, *152*; Gemini, 179; migration and, 85, 104; Paris terrorist attacks and, 85; regulation and, 214n106
Graphika, 136
Great Deplatforming, 102, 105
Green Movement, 176
Guardian, The (newspaper), 174

Hamas, 177–78
harassment: Avia Law and, 58–59, 65, 181; convergence and, *157*, 167; future issues and, 172; keywords and, *195*; messaging and, 137–38; migration and, 71, 72, 80, *81*, 86; mobilization and, 101; moderation and, 1, 8, 10; NGOs and, *185*, 186; regulation and, 45, 58
harmful content: AI-generated, 178–79; algorithms and, 115–16; American Defense League, 172; categories of, 3, 35, 50, 71–73, 156, 158, 165–66, 175, *194*, 196, 228n23; content adaptation and, 3–4, 8, 17, 19, 24, 29–30, 36–38, 41, 43, 123, 148, 164, 168, 175; convergence and, 148–58, 161–66, 196,

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

227n1, 228n23; defining, 9–11; digital resilience and, 20–29, 35–44; diverse online environment and, 4; future issues and, 168–80, 230n2; Global Internet Forum to Counter Terrorism (GIFCT) and, 38; information operations and, 173–75; keywords and, 39, 157, *194–96*; legislation of, 2, 23, 29, 47, 53, 55, 58, 60, 64–67, 167; lenient content policies and, 23–26, 44, 148, 171–73; manipulation and, 1, 42, 72, 103; messaging and, 123, 138; Meta and, 3; migration and, 70–73, 83, 85, 98; mobilization and, 101–3; moderation and, 1–19; and the Online Safety Act, 29; online thriving of, 4–8; propaganda, 9, 38–39, 42, 70, 154, 165, 173; QAnon and, 10; regulation and, 46–67; Telegram and, 10–11

hashtags, 90, 136, 140–43, *193*, 226n62

hate speech: advertisements and, 1; algorithms and, 115–16; Center for Countering Digital Hate, 172; Code of Conduct and, 59–60; convergence and, *157*, 158, 166; dangerous actors and, 17, 24, 44, 71, 80–81, 96, 98, 158, 166, 170, 185–86, 194; deplatforming and, 115–16; digital resilience and, 20, 31, 36; Facebook and, 14, 31, 54, 57, 59, 96, 101–2; future issues of, 172; Germany and, 14, 53–58; keywords and, *195*; lenient content policies and, 13, 18, 82, 98, 105, 148, 169–71; manipulation and, 1, 73, 103; measuring, 187–88, *189–92*; migration and, 72, 80, *81*, 96; militant groups and, 3–6, 11–44, 68–82, 98–103, 115, 121, 123, 146–49, 154, 161–71, 184–85, 204n2, 216n8; mobilization and, 101–5, 115–16, *117*, 121; moderation and, 1, 8, 14; nationalism and, 8, 75, *76*, 116, 157, 187, *194*; Nazism and, 53–54, 72, 75, 115; NGOs on, *185*, 186; regulation and, 50–60, 67; Southern Poverty Law Center (SPLC) and, 75–77, 95, 183–84; terrorists and, 45, 51, 58, 115, 166; 2020 elections and, 14, 18, 103, 105, 115; Twitter and, 31, 57, 59, 105, 114–17, 172, 187, 221n31; United States and, 14, 75–76, 78, 105, 116, 166, 187; white supremacists and, 50, 72, 172; YouTube and, 31, 59

Hate Speech and Disinformation Prevention and Suppression Proclamation, 67

"Hearing on Social Media Influence in the 2016 United States Elections" (Congressional hearing), 53

hostages, 1

Human Rights Watch, 64

India, 66

influence campaigns, 13–16, 19, 102, 124, 173–74

information flows, 24–26, 166

Instagram: Afghanistan and, 127; audience size of, 73; China and, 174; convergence and, 228n20, 228n24; future issues and, 174–77; messaging and, 127–28, 138; migration and, 72–73, 215n4; regulation and, 50, 53, 61; research methodology and, 184

Institute for Strategic Dialogue, 146

Insurrection Act, 100, 118

Intermediary Guidelines and Digital Media Ethics Code, 66

Internet Archive, 156, 194, 196, 228n20

Iran, 176

Iraq, 82, 84, 144, 183

Islamic State (IS): convergence and, 154, 166; Dawn of Glad Tidings, 83; digital resilience and, 31, 34, 37–41, 204n6; Europol and, 90–93; Facebook and, 1, 5–6, 19, 31, 39, 83, 85, 87, 91, 122, 144–46, 174; future issues and, 172, 174, 180; hostages and, 1; Jordanian pilot execution and, 1; Mapping Militants Project and, 183; messaging and, 122, 128, 144–46; migration and, 5, 69–70, 82–93, 96; moderation and, 1, 5–6, 13, 18–19; regulation and, 49, 58; Telegram and, 5, 31, 82, 89–93, 122; Twitter and, 1, 5–6, 18–19, 31, 37, 82–93, 96, 122, 128, 146, 172, 217n25; video and, 1, 39, 83, 204n6

Israel, 177–78, 184

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

Jews, 53, 172, 184, 187
jihadi groups, 37, 50, 69, 90, 166, 172, 183
Jordanian pilot murder, 1
Jourová, Věra, 59
JustPaste.it, 91

Kaplan, Anna, 51
keywords: authenticity, *196*; bullying, *195*; convergence, 156, *157*, 164; digital resilience, 39; disinformation, *196*; harassment, *195*; harmful content, 39, 157, *194–96*; hate speech, *194–95*; manipulation, *196*; messaging, 130, 224n35, 226n72; mobilization, 113, 119, 221n34, 222n62; terrorists, *194–95*; violence, *194–96*; white supremacists and, *157*

League of Women Voters, 207n7
Lebanon, 176
legal issues: child sexual abuse material (CSAM), 173–76; Code of Conduct and, 59–60; convergence and, 166–67; *FCC v. League of Women Voters*, 207n7; migration and, 71, 96; mobilization and, 103; moderation and, 1, 3; personal liability, 48; regulation and, 46, 48–60, *63*, 64, 213n88; research methodology and, 182
legislation: accountability and, 51–52; Avia Law, 58–59; Biden and, 51–52; Christchurch Call to Action, 15, 28, 44, 151–53, 170; democratic, 23, 45, 52, 67, 167; digital resilience and, 23, 29, 46; European, 53, 58–62; FCC and, 47; future issues for, 177; Global Internet Forum to Counter Terrorism (GIFCT), 15, 38, 153–54, 158, 162, 165–67, 227n13; global view on, 64–67; of harmful content, 2, 23, 29, 47, 53, 55, 58, 60, 64–67, 167; moderation and, 2, 5; Online Safety Act, 29, 66; research methodology on, 64–67; Russian, 167; Telecommunications Act, 47–52; Trump and, 52
lenient content policies: convergence and, 148–50, 160–61, 166; digital resilience and, 23–26, 43–44; future issues and, 169–74, 177; harmful content and, 23–26, 44, 148, 171–73; hate speech and, 13, 18, 82, 98, 105, 148, 169–71; messaging and, 122; migration and, 70, 73, 82, 85–89, 97–98; militant groups and, 13, 18, 82, 99, 148, 169–71; mobilization and, 99, 102–4, 115, 117, 121; moderation and, 5, 13, 18; Musk and, 171–73; research methodology and, 186; Telegram and, 229n30
liberal values: democracy and, 45, 67–68; freedom of expression, 17, 21, 23, 47, 53, 56–61, 67, 85; immigration and, 188; moderation and, 17; regulation and, 45, 67, *68*
Libya, 176
LinkedIn, 61, 184, 214n106, 215n4, 228n24

Macron, Emmanuel, 28, 59, 151
manipulation: artificial intelligence (AI) and, 1; of audiences, 72–73; convergence and, *157*, 167; elections and, 42, 72, 103–4; harmful content and, 1, 42, 72, 103; hate speech and, 1, 73, 103; keywords and, *196*; messaging and, 125, 128; migration and, 72–73; mobilization and, 103–4; research methodology and, *196*; 2020 elections and, 72, 103
Mapping Militants Project (MMP), 75–77, 183
Mastodon, 186, 220n20, 234n1
Matrix, 87
McInnes, Gavin, 94–96
Merkel, Angela, 54
messaging: abuse and, 138; Al Qaeda and, 128; Al-Shabaab and, 128; authenticity and, 18, 22, 24, 26, 29–31, 43, 117, 123, 147, 169; bans and, 123, 128, 133, 143–45; censorship and, 123; conspiracy and, 136–39, 143; content adaptation and, 35–37, 123–45; convergence and, 148–50, 155, 161–62; Conversations, 87; deplatforming and, 131, *132*, 139; digital resilience and, 21, 29, 33–38, 41–44; Discord, 87; distribution of, 37–38; divergence and, 35–41; Europe and, 136; extremists and, 123, 128,

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

135, 137, 142, 147; Facebook, 122–27, 133, 136–39, 143–46; future issues and, 169–70, 177; Gab, 139–42, 226n64; harassment and, 137–38; harmful content and, 123, 138; Instagram, 127–28, 138; Islamic State (IS) and, 122, 128, 144–46; keywords and, 130, 224n35, 226n72; lenient content policies and, 122; manipulation and, 125, 128; Matrix, 87; migration and, 6, 13–19, 82–83, 87; militant groups and, 123–28, 134, 144–47; mobilization and, 101, 115–21; moderation threshold and, 123–25, 131, 134–35, 138–39, 143–44; Parler, 139; Pinterest, 138; propaganda and, 122, 125–26, 129, 143–47; QAnon, 124, 134–43; racism and, 139; research methodology, *192–93*; susceptibility of audience and, 18, 33, 35, 37, 101, 121, 170; suspensions and, 123, 129–35, 138, 141, 145, 222n1; Taliban and, 123–34, 143, 145–46, 223n19, 224nn35–36, 226n72; technology and, 124–25, 135, 137; Telegram, 122, 128; terrorists and, 127–28, 137, 144; threshold evasion and, 38–41; trackers and, 129; Twitter, 122–46, 222n1, 224n35, 226n64, 226n72; unaffiliated accounts and, 144–46; United States and, 126, 136–38, 144; violence and, 122, 127–34, 137–38, 144–46, 224n35, 226n72; YouTube, 122–28, 136–38

Meta: AI leak of, 178; Breton letter and, 178; convergence and, *152*; CSAM and, 175; future issues and, 174–75, 178; harmful content removal and, 3; moderation and, 2–3; regulation and, 50

MeWe, 14

Microsoft: Code of Conduct and, 59; convergence and, *152*, 153; future issues and, 179; Global Internet Forum to Counter Terrorism (GIFCT) and, 153; migration and, 85; Paris terrorist attacks and, 85

Middle East, 1, *76*, 84, 89

migration: abuse and, 86, 91; advertisements and, 85, 90, 95; Al Qaeda and, 31; Al-Shabaab and, 78; authenticity and, 70, 73, 75, 79, 82, 91–93, 98; automation and, 83–86; bans and, 69, *72*, 73–74, 79, 82, 93, 96–97; Big Tech and, 111–15; bullying and, 71, *72*, 80, *81*; collaboration and, 89, 91; dangerous actors and, 71, 80–81, 96, 98; deplatforming and, 69–70, 78, 82; digital resilience and, 69, 93; Discord and, 215n4; divergence and, 30–33, 69; Europe and, 78, 84, 90, 94; Europol and, 90–93; extremists and, 5, 7, 11, 13, 17, 19, 21, 30, 35, 44, 69–71, 74–77, 81–87, 100, 104, 123, 148–50, 161, 170; Facebook and, 70–78, 83–98, 215n4; freedom of expression and, 67, 85; Gab and, 73–79, 96–98, 215n4, 218n2, 220nn20–21, 220n24; Google and, 85, 104; harassment and, 71, *72*, 80, *81*, 86; harmful content and, 70–73, 83, 85, 98; hate speech and, 72, 80, *81*, 96; Instagram and, 72–73, 215n4; Islamic State (IS) and, 5, 69–70, 82–93, 96; legal issues and, 71, 96; lenient content policies and, 70, 73, 82, 85–89, 97–98; LinkedIn and, 215n4; locations of, 74–82; manipulation and, 72–73; messaging and, 82–83, 87; Microsoft and, 85; militant groups and, 69–82, 87, 93, 98; misinformation and, 72, *81*; moderation threshold and, 70–73, *74*, 79, *80*, 86–87, 91–93; Nazism and, *72*, 75–78; Parler and, 73–77, 96–98, 215n4; Pinterest and, 79, 87–88, 215n4; propaganda and, 69–70, 82–91, 95; Proud Boys and, 18, 69–71, 82, 93–98; racism and, 76; recruitment and, 69, 82, 84–85, 93–99; Reddit and, 75, 101, 215n4; suspensions and, 69, 75, 78–79, 86–87, 89, 95; technology and, 70–71, 82–86, 93, 95, 98; Telegram and, 75, *78*, 79, 82, 88–93, 97–98, 215n4; terrorists and, 71–72, 85–86, 89–91, 115; TikTok and, 73, *78*, 79, 215n4; Twitter and, 69–78, 82–97, 215n4; United States and, 75–78, 84–85, 93–95; video and, 70, 83, 89–90, 95; violence and, 71–72, 75, 80–81, 84–86, 94–98; YouTube and, 70–78, 83–87, *88*, 94, 215n4

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

militant groups: content adaptation and, 17–18, 24, 29–30, 36–38, 43, 123, 134, 148, 164, 168; convergence and, 148–49, 154, 161–67; dangerous actors and, 17–18, 24, 32, 44, 71, 80–81, 166, 170; digital resilience and, 20–25, 29–44; and future issues, 168–71; hate speech and, 3–6, 11, 13–19; Islamic, 13, 18, 39, 59, 82, 128, 146; lenient content policies and, 13, 18, 82, 99, 148, 169–71; Mapping Militants Project (MMP), 75–77, 183; messaging and, 123–28, 134, 144–47; migration and, 69–82, 87, 93, 98; mobilization and, 99–103, 115, 121; moderation and, 3–6, 11–18; regulation and, 59, 68; research methodology and, 183–85; unaffiliated accounts and, 144–46; violence and, 3, 11, 14, 39, 75, 80–81, 134, 165

misinformation: convergence and, 157; digital resilience and, 22, 42; disinformation, 19, 53, 67, *157*, 172–74, 181, *196*; future issues and, 172, 179; influence operations and, 13–16, 19, 124, 173–74; keywords for, *196*; migration and, 72, *81*; mobilization and, 103–5; moderation and, 1, 3, 9–10, 14; NGO accounts on, *185*, 186; regulation and, 45, 51, 53; 2020 elections and, 14, 72, 103, 105

mobilization: abuse and, 99, 104, 115; automation and, 108; bans and, 100–4, 111, 116; censorship and, 99–103, 107–8, 111–14, 121, 219n14; conspiracy and, 107–8; deplatforming and, 100–21; divergence and, 33–35; extremists and, 100–2, 104, 108–10, 115–21; Facebook and, 99–104, 117–18, 121; Gab and, 99–100, 105–21; harassment and, 101; harmful content and, 101–3; hate speech and, 101–5, 115–16, *117*, 121; influencers and, 13–16, 102; keywords and, 113, 119, 221n34, 222n62; legal issues and, 103; lenient content policies and, 99, 102–4, 115, 117, 121; manipulation and, 103–4; messaging and, 101, 115–21; militant groups and, 99–103, 115, 121; misinformation and, 103–5; moderation threshold and, 105; Oath Keepers and, 18, 99–100, 110, 115–21; platform usage patterns and, 110–11; propaganda and, 100–3, 114, 121; QAnon and, 108; recruitment and, 3, 6, 11, 13, 18, 33, 49, 85, 93, 98, 101–3, 117–21, 150, 170; suspensions and, 99–105, 108–20, 221n33; technology and, 103–4, 113, 219n14; 2020 elections and, 99, 103–8, 115, 118; Twitter and, 99–100, 104–21, 220n21, 220n24, 221n29, 221n31; United States and, 99, 105, 114–18; user experiences with, 104–10; video and, 99, 114, 118; violence and, 100, 108–10; YouTube and, 99

moderation: abuse and, 1, 15; algorithms and, 6, 39, 144, 152, 213n88, 228n23; authenticity and, 15, 18, 24, 26, 29–33, 36, 70, 73, 79, 82, 91, 98, 117, 123, 147, 149, 163, 169–72; automated, 6, 27–28, 39, 57, 85–86, 125, 138, 144, 228n20; bans and, 5–7, 13, 15, 17; censorship and, 12–19; centralized, 15–16, 19, 164–67; collaboration and, 3, 12, 15; collateral damage by, 15, 164; conspiracy and, 6, 10; content adaptation and, 3–8, 11, 16–19, 148, 150, 162, 164; convergence in, 148–67; cycles of, 41–43; dangerous actors and, 7, 12; democracy and, 9; deplatforming and, 18; different standards for, 169–71; digital resilience and, 4, 16; divergence and, 4, 8–9; extremists and, 1–19, 148–57, 161–67; Facebook and, 1–2, 5–6, 10, 14, 16, 19; freedom of expression and, 17; future of, 168–80; Gab and, 2–3, 10, 14, 19; Germany and, 14, 16; greater, 84–86; harassment and, 1, 8, 10; harmful content and, 1–19, *194–96*; hate speech and, 1, 8, 14; information flows and, 24–26, 166; Islamic State (IS) and, 1, 5–6, 13, 18–19; legislation on, 2, 5; lenient content policies and, 5, 13, 18; messaging and, 6, 13–19; Meta and, 2–3; militant groups and, 3–6, 11–18; misinformation and, 1, 3, 9–10, 14; mobilization

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

and, 6–7, 69–98; Parler and, 14; PCVE and, 8; propaganda and, 4, 6, 9–13, 18; Proud Boys and, 18; QAnon and, 6, 10, 18; recruitment and, 3, 6, 11, 13, 18; sanctions and, 3, 23, 102, 174, 227n13; self, 28, 46, 48–51, 55, 64–67, 100; suspensions, 3, 6; takedowns, 3, 72, 87, 91, 123, 139, 165; Taliban and, 18, 127–28; technology and, 1–7, 10–11, 15–17; Telegram and, 1, 3, 5, 16; terrorists and, 2–11, 14–15; TikTok and, 3, 16; transparency and, 2, 19; 2020 elections and, 103–4; Twitter and, 1–2, 5–6, 10, 14–19; United States and, 9, 14, 16, 46; user experiences with, 104–10; violence and, 1, 3, 8, 11, 14–17; white supremacists and, *72*; YouTube and, 1–2, 10, 14, 16
moderation threshold: alignment in, 148–50; changing policies and, 27–30; convergence and, 148–50, 155–67; digital resilience and, 21–40, 43; evasion of, 38–41, 143–46; messaging and, 123–25, 131, 134–35, 138–39, 143–44; migration and, 70–73, *74*, 79, *80*, 86–87, 91–93; mobilization and, 105; research methodology and, 196; single, 26–30
Mohammad, Saif, 112
monetization, 34, 102
Mujahid, Zabihullah, 129
murder: Jordanian pilot and, 1; terrorists and, 1–2, 27, 51, 177; glorification of violence and, 128
Musk, Elon, 171–73, 231n9, 233n42
Muslim Brotherhood, 72–73

National Center for Missing and Exploited Children, 175
nationalism: extremists and, 8, 75, *157*, *194*; hate speech and, 8, 75, *76*, 116, 157, 187, *194*; research methodology and, 187–88, *194*; white supremacists and, 75, 116, 187–88
Nazism: Atomwaffen Division and, 78; convergence and, *157*; Germany and, 53–54; hate speech and, 53–54, 72, 75, 115; migration and, *72*, 75–78; research methodology and, 184, *194*; SPLC and, 75, *76*; Twitter and, 115
Network Enforcement Act (NetzDG): controversy over, 56; Criminal Code and, 56; obligations of, 55–56; platform focus of, 56–57; Germany and, 29, 53–58, 62, 65, 181; rules of, 55–56
New Zealand: Ardern and, 28, 151; Christchurch terrorist attack, 2, 15, 27–28, 44, 60, 151–53, 166, 170; Facebook and, 2, 27–28, 60, 151–53; Films, Videos and Publications Classification Amendment Bill, 65; Global Internet Forum to Counter Terrorism (GIFCT), 15, *166*; Reddit and, 151
nongovernmental organizations (NGOs), 82, 184–86

Oath Keepers: advertisements and, 119; Biden and, 99–100, 118–19; “Defeat the Coup” campaign and, 18, 100, 118; Gab and, 99–100, 110, 115–21, *189*, 222n54, 222n58, 222nn61–62; mobilization and, 18, 99–100, 110, 115–21; research methodology and, *189–90*, *192*; “Stop the Steal” campaign and, 18, 100, 118; Trump and, 100, 118–19; US Capitol attack and, 18, 100, 119
Obama, Barack, 49
Online Regulation Series, 64
Online Safety Act, 29, 65, *66*
OpenAI, 179
Orange Revolution, 176

Pakistan, 67
Parler: convergence and, 160, *161*, 167; digital resilience and, 31; messaging and, 139; migration and, 73–77, 96–98, 215n4; moderation and, 14; Proud Boys and, 96–98; research methodology and, 184
Pashto, 129–30, 145–46, *192–93*
Patriot Front, *78*, 183
patriots, 114, 119, 135

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

pCloud, 87, 87–88
peace, 91, 184
PeerTube, 73
Pinterest: convergence and, 160, *161*, 228n24; messaging and, 138; migration and, 79, 87–88, 215n4; policy alignment and, 160; regulation and, 61, 214n106; research methodology, 184
polarization, 1, 9, 14
pornography, 98
preventing or countering violent extremism (PCVE), 8
"Promoting Our European Way of Life" (European Commission), 60
propaganda: convergence and, 154, 162–67; Dawn of Glad Tidings, 83; digital resilience and, 24, 37–42, 204n6; Facebook and, 39–40, 50, 83, 85, 100, 122, 125, 144–45, 165, 174; future issues and, 168–74; information operations and, 173–75; messaging and, 122, 125–26, 129, 143–47; migration and, 69–70, 82–91, 95; mobilization and, 100–3, 114, 121; moderation and, 4, 6, 9–13, 18; reaching the masses, 83–84; regulation and, 45, 49–50, 54, 56; susceptibility of audience and, 18, 33, 35, 37, 101, 121, 170; terrorists and, 4, 6, 9, 38, 45, 50, 85–86, 91, 144, 154; TikTok and, 174; Twitter and, 37, 40, 50, 83–86, 91, 100, 114, 122, 125; unaffiliated accounts and, 144–46; video and, 1, 39, 50, 83, 204n6; YouTube and, 40, 50, 83, 85, 122, 174
Proud Boys: Facebook and, 94–97; First Degree of, 95; future issues and, 180; Gab and, 96–98; McInnes and, 94–96; migration and, 18, 69–71, 82, 93–98; moderation and, 18; Parler and, 96–98; platform policy changes over, 95–96; recruitment and, 93–98; research methodology and, 184; Southern Poverty Law Center (SPLC) and, 95; Telegram and, 97; Trump and, 97–98; Twitter and, 96–97; "Unite the Right" rally, 96; violence of, 95; Western chauvinism and, 94
QAnon: advertisements and, 136; conspiracy and, 6, 10, 34, 40, 108, 136–39, 143; content adaptation and, 134–43; digital resilience and, 34, 40; Facebook and, 6, 10, 34, 40, 136–39, 143; FBI and, 137; future issues and, 180; Gab and, 10, 108–42, *193*; Graphika and, 136; harmful content and, 10; mainstream media and, 135–36; messaging and, 124, 134–43; mobilization and, 108; moderation and, 6, 10, 18; Q Clearance Patriot, 135; Reddit and, 135–37; #SaveTheChildren and, 143; Trump and, 135; Twitter and, 6, 10, 18, 34, 40, 108, 136–43, *193*, 221n30, 226n64; US Capitol attack and, 6, 34; YouTube and, 10, 40, 136, 136–38

racism: toward African Americans, 51, 172, 187; toward Asians, 187; and Buffalo supermarket attack, 51; convergence and, 157; toward Jews, 53, 172, 184, 187; messaging and, 139; migration and, 76; regulation and, 51–55; research methodology and, *195*; Southern Poverty Law Center (SPLC) and, 75–77, 95, 183–84; xenophobia and, 53–54, 178
radicalization: Christchurch Call to Action and, 153; deplatforming and, 180; extremists and, 13, 33, 49, 51, 84–85, 177, 180; online communities and, 13, 33, 49, 51, 85; Oath Keepers and, 118; propaganda and, 84; TikTok and, 177
Ramadan, 37–38
recruitment: Al Qaeda and, 13; convergence and, 150; digital resilience and, 22, 33, 206n20; extremists and, 6, 11, 13, 18, 33, 85, 118; future issues and, 170; migration and, 69, 82, 84–85, 93–99; mobilization and, 3, 6, 11, 13, 18, 33, 49, 85, 93, 98, 101–3, 117–21, 150, 170; moderation and, 3, 6, 11, 13, 18, 33, 49, 85, 93, 98, 102, 121, 150, 170; Proud Boys and, 93–98; radicalization and, 13, 33, 49, 51, 84–85, 180; regulation and, 3, 49
Reddit: China and, 174; convergence and, 228n24; migration and, 75, 215n4; mobili-

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

zation and, 101; New Zealand massacre and, 151; QAnon and, 135–37; research methodology and, 184
reform, 51–52
regression analysis, 79, *80*, 91, *92*, 162, 226n64
regulation: abuse and, 54; algorithms and, 52; Al Qaeda and, 50; Avia Law, 58–59, 65, 181; censorship, 46, 48, 52, 56; Central Office for the Fight Against Crime Related to Information and Communication Technology, 58; Code of Conduct, 59–60; Code of Practice on Disinformation, 172; collaboration and, 59–60; Communications Act, 47; dataset for, 181–83; democracy and, 45, 52, 65, 67–68; deplatforming and, 52; digital resilience and, 46, 62; Digital Services Act, 2; Europe and, 53–65; extremists and, 46, 49–54, 57, 62, 67–68; Facebook and, 49–50, 53–61, 214n106; FCC and, 47; Films, Videos and Publications Classification Amendment Bill, 65; freedom of expression and, 47, 53, 56–61, 67; Global Internet Forum to Counter Terrorism (GIFCT), 15, 38, 153–54, 158, 162, 165–67, 227n13; global view on, 64–67; harassment and, 45, 58; harmful content and, 46–67; hate speech and, 50–60, 67; Instagram and, 50, 53, 61; Insurrection Act, 100, 118; Islamic State (IS) and, 49, 58; large platforms and, 52–53; legal issues and, 46, 48–60, *63*, 64, 213n88; legislation, 2, 5 (*see also* legislation); liberal values and, 45, 67, *68*; limiting, 47–49; Meta and, 50; militant groups and, 59, 68; minimal approach to, 46–53; misinformation and, 45, 51, 53; Network Enforcement Act (NetzDG), 29, 53–58, 62, 65, 181; "one-hour rule" and, 59–60; Online Safety Act, 29, 66; Pinterest and, 61, 214n106; propaganda and, 45, 49–50, 54, 56; racism and, 51–55; recent trends in, 227n1; recruitment and, 3, 49; research methodology on, 64–67; sanctions and, 3, 23, 102, 174, 227n13; Section 230 and, 52; self, 48–49; strategic behavior in, 21–24; takedowns, 3, 72, 87, 91, 123, 139, 165; technology and, 45–64, 67, 207n7; Telecommunications Act, 47–52; TERREG and, 60–61; terrorists and, 45, 49–51, 55, 58–61, 64; TikTok and, 61; transparency and, 57, 62, *63*; Twitter and, 49–50, 53–61; United States and, 45–53; video and, 50, 65; violence and, 49–51, 56–57, 62; YouTube and, 49–50, 53, 55–61
Republicans, 9, 52, 99
research methodology: abuse and, 187; bans and, 185; dangerous actors, *185*, *194*; databases, 181–86; deplatforming, 108–10, *191–92*; Facebook and, 184; future issues and, 177–80; Instagram and, 184; keywords, 113 (*see also* keywords); legislation and, 64–67; lenient content policies, 186; manipulation and, *196*; militant groups and, 183–85; nationalism and, 187–88, *194*; Nazism and, 184, *194*; NGOs and, 184–86; Oath Keepers and, *189–90*, *192*; Pinterest and, 184; platform usage patterns and, 110–11; racism and, *195*; Reddit and, 184; regression analysis, 79, *80*, 91, *92*, 162, 226n64; regulation and, 64–67; security, 181, 184; suspensions, *189–92*, *197*; Taliban and, 224n35, 226n72; technology and, 196; Telegram and, 184; TikTok and, 184; Twitter and, 184–87, *189–93*, 218n2, 220n21, 220n24, 221n29, 221n31; United States and, 187; Wayback Machine and, 156, 194, 196, 228n20; white supremacists and, 188, *194*
Roberts, Margaret, 34, 39, 103
Rocket.Chat, 91, 160
Rumble, 3, 160, *161*
Russia, 14, 16, 174

sanctions, 3, 23, 102, 174, 227n13
Saudi Arabia, 167, 176
#SaveTheChildren, 143
scandals, 42–43

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

Section 230, 47–52
security: Obama administration and, 49; operational, 32; propaganda and, 49; research methodology, 181, 184; Stanford Center for International Security and Cooperation, 75; US officials and, 49, 85–86
Security Union, 60
sexism, *157*
sex trafficking, 135–36, 143
sexual orientation, 72, 115, 187, *195*, 230n7
sexual violence, 58
Shinas, Margaritis, 60
Singapore, 65, 182
skinheads, *76*
Snapchat, 61, 87–89, 175
Southern Poverty Law Center (SPLC), 75–77, 95, 183–84
spam, 98, 128, *157*, 174, *196*
Specially Designated Global Terrorists (SDGTs), 127
Spiegel, Der (newspaper), 174
Stanford's Center for International Security and Cooperation, 75
"Stop the Steal" campaign, 18, 100, 118
susceptibility, 18, 33, 35, 37, 101, 121, 170
suspensions: convergence and, 162, *163*, 166; digital resilience and, 31–34, 43; Facebook, 34, 99, 125, 133, 166, 222n1; future issues and, 170, 172; Gab, 105, 109–12, 116, 119, *189*, *191*; messaging and, 123, 129–35, 138, 141, 145, 222n1; migration and, 69, 75, 78–79, 86–87, 89, 95; mobilization and, 99, 101–5, 108–20, 221n33; moderation and, 3, 6; research methodology and, *189–92*, *197*; Twitter, 34, 95, 99, 105, 108–14, 117–20, 125, 129–34, 138, 145, *191–92*, 222n1; user's anger over, *191*; YouTube, 125
Syria, 54, 82, 84, 165, 176, 216n15, 229n30

takedowns, 3, 72, 87, 91, 123, 139, 165
Taki magazine, 94
Taliban: Afghanistan and, 126–27; Biden and, 127; content adaptation and, 18, 123–25, 128, 131, 134, 143, 145; digital resilience and, 40; early adoption of online media by, 125–26; Facebook and, 125–27; future issues and, 180; Instagram and, 127; Mapping Militants Project and, 183; Mazar-i-Sharif and, 126–27; messaging and, 123–34, 143, 145–46, 223n19, 224nn35–36, 226n72; moderation and, 18, 127–28; online amplification by, 126–27; research methodology on, 224n35, 226n72; Specially Designated Global Terrorists (SDGTs) list and, 127; Telegram and, 128; Twitter and, 18, 40, 123–34, 143–46, *192–93*, 224n35, 226n72; unaffiliated accounts and, 145–46; video and, 124, 128, 146; violence and, 128–34, 145–46, 224n35, 226n72; WhatsApp and, 127; YouTube and, 125–28
TamTam, 70
Tech Against Terrorism, 64, 154
technology: advertisements and, 22; artificial intelligence (AI), 1, 39, 49, 174, 178–80; automation, 6, 27–28 (*see also* automation); Big Tech, 51–53, 97, 111–15, *189–91*; convergence and, 150–55, 160, 164–67, 196, 227n13; digital resilience and, 20–27, 35–42, 204n2; and future issues, 168–80, 230n2; harassment and, 1 (*see also* harassment); liberation, 176–77; messaging and, 124–25, 135, 137; migration and, 70–71, 82–86, 93, 95, 98; mobilization and, 103–4, 113, 219n14; moderation and, 1–7, 10–11, 15–17; regulation and, 45–64, 67, 207n7; research methodology and, 196; Section 230 and, 52; United States and, 45–46, 53, 85
Telecommunications Act, 47–52
Telegram: audience size of, 73; bans, 5, 36, 82, 93, 97; content adaptation and, 5, 36, 229n30; convergence and, 160, *161*, 229n30; CSAM and, 175; digital resilience and, 23, 28–29, 31, 36; Durov, 89; Europol and, 90–93; future issues and, 174–77; harmful content and, 10–11; Islamic State

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

(IS) and, 5, 31, 82, 89–93, 122; lenient content policies and, 229n30; messaging and, 122, 128; migration and, 75, *78*, 79, 82, 88–93, 97–98, 215n4; moderation and, 3, 5, 16; Proud Boys and, 97; research methodology and, 184; Taliban and, 128; Trump and, 97–98
TERREG, 60–61
terrorists: Arabic language and, 129–30, 144, 154, *192–93*, 228n21; Belgium and, 14; Buffalo supermarket attack, 51; Christchurch mosque attack, 2, 15, 27–28, 44, 60, 151–53, 166, 170; convergence and, 147, 151–54, 166–67; Dari language and, 129–30, 145–46, *192–93*; digital resilience and, 27–28, 38–42; domestic, 51, 137; France and, 14, 28, 58–59, 85; future issues and, 177; Germany and, 14, 55; hate speech and, 45, 51, 58, 115, 166; jihadi groups, 37, 50, 69, 90, 166, 172, 183; keywords and, *194–95*; messaging and, 127–28, 137, 144; migration and, 71–72, 85–86, 89–91, 115; moderation and, 2–11, 14–15; murder and, 1–2, 27, 51, 177; Paris attacks and, 49, 58, 85, 151; Pashto language and, 129–30, 145–46, *192–93*; propaganda and, 4, 6, 9, 38, 45, 50, 85–86, 91, 144, 154; regulation and, 45, 49–51, 55, 58–61, 64; San Bernardino, California, 49; Specially Designated Global Terrorists (SDGTs), 127; white supremacists and, 27, 50. *See also* specific organization
Texas, 52–53, 210n44
Thompson, Bennie, 28
TikTok: AI-generated messages on, 178; audience size of, 73; Breton letter and, 178; China and, 174; convergence and, 160, *161*, 228n24; CSAM and, 175; digital resilience and, 36; future issues and, 174–78; harmful content removal and, 3; migration and, 73, *78*, 79, 215n4; moderation and, 3, 16; policy alignment and, 160; propaganda on, 174; regulation and, 61; research methodology and, 184; video and, 3, 175, 178
"Together against Hate Messages" (German task force), 55
Top4top, 91
trackers, 129
transparency: convergence and, 153, 227n13; digital resilience and, 29, 38–39; moderation and, 2, 19; regulation and, 57, 62, *63*
Trump, Donald: bias and, 52, 107, 114; "Defeat the Coup" campaign and, 18, 100, 118; election issues and, 97–100, 107–8, 114–15, 118–19; Facebook and, 221n1; Insurrection Act and, 100, 118; legislation and, 52; Oath Keepers and, 100, 118–19; Proud Boys and, 97–98; QAnon and, 135; Section 230 and, 52; "Stop the Steal" campaign and, 18, 100, 118; Telegram and, 97–98; 2020 elections and, 97, 99, 107–8, 118; Twitter and, 99, 107–8, 114, 222n1; US Capitol attack and, 97–100, 107–8, 118, 222n1; white supremacists and, 172
Tumblr, 87, *88*, 160, *161*
Tunisia, 176
Turney, Peter, 112
2020 elections: Gab and, 14, 105, 115, 118, 121; hate speech and, 14, 18, 103, 105, 115; manipulation and, 72, 103; misinformation and, 14, 72, 103, 105; mobilization and, 99, 103–8, 115, 118; moderation and, 103–4; Trump and, 97, 99, 107–8, 118; Twitter and, 99, 105, 107, 115; US Capitol attack and, 14, 97–100, 107–8, 118, 222n1
Twitter: Al Qaeda and, 31, 128; audience size of, 73; bans and, 5–6, 40, 82, 93, 96, 100, 111, 116, 128, 133, 143; censorship and, 111–15; Code of Conduct and, 59; content adaptation and, 5–6, 18–19, 123, 125, 131, 134, 140–45, 229n30; convergence and, 152–53, 160–61, 228n24, 229n30; Congressman Thompson letter to, 28; data collection on, 186–87, *189–93*; Dawn of Glad Tidings and, 83; digital resilience and, 28, 31, 34, 37, 40; fall of, 171–73; Global Internet Forum to Counter Terrorism (GIFCT) and, 153; hate speech and, 31, 57, 59, 105,

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

Twitter (*continued*)
114–17, 172, 187, 221n31; Islamic State (IS) and, 1, 5–6, 18–19, 31, 37, 82–93, 96, 122, 128, 146, 172, 217n25; messaging and, 122–46, 222n1, 224n35, 226n64, 226n72; migration and, 69–78, 82–97, 215n4; mobilization and, 99–100, 104–21, 220n21, 220n24, 221n29, 221n31; moderation and, 1–2, 5–6, 10, 14–19; Musk and, 171–73; Nazism and, 115; Paris terrorist attacks and, 85; platform usage patterns and, 110–11; policy alignment and, 160; propaganda on, 37, 40, 50, 83–86, 91, 100, 114, 122, 125; Proud Boys and, 95–97; QAnon and, 6, 10, 18, 34, 40, 108, 136–43, *193*, 221n30, 226n64; regulation and, 49–50, 53–61; research methodology and, 184–87, *189–93*, 218n2, 220n21, 220n24, 221n29, 221n31; suspensions and, 34, 95, 99, 105, 108–14, 117–20, 125, 129–34, 138, 145, *191–92*, 222n1; Taliban and, 18, 40, 123–34, 143–46, *192–93*, 224n35, 226n72; Trump and, 99, 107–8, 114, 222n1; 2020 elections and, 99, 105, 107, 115; US Capitol attack and, 6, 34, 105, 108, 222n1. *See also* X
Twitter Rules, 71
Twitter Transparency, 200n14

Ukraine, 16, 174, 176
United Kingdom, *152*, 183; Britain First, 31; greater moderation and, 84–85; Online Safety Act, 29, 66
United States: Buffalo supermarket attack, 51; California, 49, 51; Capitol attack, 6, 14, 18, 34, 51, 100–5, 108, *109*, 119, 139, 222n1; Communications Act, 47; Congressman Thompson letter, 28; convergence and, *152*, 166; Democrats, 9, 52; digital resilience and, 28, 40; Florida, 52–53, 210n44; hate speech and, 14, 75–76, 78, 105, 116, 166, 187; large platforms and, 52–53; liberal democracy and, 45; messaging and, 126, 136–38, 144; migration and, 75–78, 84–85, 93–95; mobilization and, 99, 105, 114–18; moderation and, 9, 14, 16, 46; neo-Nazis and, 78; New York, 51, 96; polarization poll in, 9; regulation and, 45–53; Republicans, 9, 52, 99; research methodology and, 187; Russia and, 14, 16; San Bernardino attack, 49; security issues and, 49, 85–86; Specially Designated Global Terrorists (SDGTs) list, 127; technology and, 45–46, 53, 85; Telecommunications Act, 47–52; Texas, 52–53, 210n44
"Unite the Right" rally, 96
Urdu, 129–30, *192–93*
US Capitol attack: "Defeat the Coup" campaign and, 18, 100, 118; Facebook and, 6, 34, 222n1; Oath Keepers and, 18, 100, 119; QAnon and, 6, 34; "Stop the Steal" campaign and, 18, 100, 118; Trump and, 97–100, 107–8, 118, 222n1; Twitter and, 6, 34, 105, 108, 222n1; violence and, 6, 14, 18, 34, 51, 100–5, 108, *109*, 119, 139, 222n1
user-generated content (UGC), 4, 22, 49, 71, 137, 169

very large online platforms (VLOPs), 61
video: convergence and, 151; digital resilience and, 22, 39; Facebook and, 1, 39, 50, 83, 95, 136, 151, 175; Films, Videos and Publications Classification Amendment Bill, 65; future issues and, 175, 178; Islamic State (IS) and, 1, 39, 83, 204n6; migration and, 70, 83, 89–90, 95; mobilization and, 99, 114, 118; propaganda, 1, 39, 50, 83, 204n6; regulation of, 50, 65; research methodology and, 182; Taliban and, 124, 128, 146; TikTok and, 3, 175, 178; violent, 1, 39, 50, 95, 128, 178; YouTube and, 1, 50, 83, 125, 128, 136, 138, 151, 175, 178
violence: Breton letter and, 178; convergence and, 151, 153, 157, 165; digital resilience and, 28, 39, 42, 204n6; extremists and, 1, 3, 8, 11, 14–17, 28, 42, 49–51, 57, 62, 81, 85–86, 108–9, 128, 137, 151–53, 157, 165, *194*; future issues and, 177–78; glorification of, 80, *195*; gun, 51; Hamas and, 177–78; keywords

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

for, *194–96*; messaging and, 122, 127–34, 137–38, 144–46, 224n35, 226n72; migration and, 71–72, 75, 80–81, 84–86, 94–98; militant groups and, 3, 11, 14, 39, 75, 80–81, 134, 165; mobilization and, 100, 108–10; moderation and, 1, 3, 8, 11, 14–17; murder, 1–2, 27, 51, 128, 177; NGO accounts of, *185*; online inspiration of, 1; Proud Boys and, 95; regulation and, 49–51, 56–57, 62; sexual, 58; shifting discourse on, 131–34; social unrest and, 23; Taliban and, 128–34, 145–46, 224n35, 226n72; US Capitol attack, 6, 14, 18, 34, 51, 100–5, 108, *109*, 119, 139, 222n1; videos of, 1, 39, 50, 95, 128, 178
Virkkunen, Henna, 61

Walk, Erin, 229n30
Wall Street Journal, 52
Wayback Machine, 156, 194, 196, 228n20
"whack-a-mole" problem, 69–70, 82
WhatsApp, 127
white supremacists: Christchurch terrorist attack and, 2, 15, 27–28, 44, 60, 151–53, 166, 170; Facebook and, 27; hate speech and, 50, *72*, 172; jihadi groups and, 50; keywords and, *157*; moderation and, *72*; nationalism and, 75, 116, 187–88; New Zealand and, 27–28; research methodology and, 188, *194*; Telecommunications Act and, 51; terrorists and, 27, 50; Trump and, 172
Wojcicki, Susan, 57
WTVideo, 70
Wyden, Ron, 48

X: Breton letter and, 178; China and, 174; Europe and, 172–73; future issues of, 171–74, 177–78; Musk and, 171–73, 231n9, 233n42; new policy of, 231n12; rise of, 171–74
xenophobia, 53–54, 178
X Transparency, 200n14

YouTube: Afghanistan and, 127; AI-generated messages on, 178; Breton letter and, 178; China and, 174; Code of Conduct and, 59; Community Guidelines, 71; Congressman Thompson letter to, 28; content adaptation and, 125, 174; convergence and, 151–53, 160, *161*, 228n24; CSAM and, 175; digital resilience and, 28, 31, 37, 40; future issues and, 168, 174–78; hate speech and, 31, 59; Jordanian pilot murder and, 1; messaging and, 122–28, 136–38; migration and, 70–78, 83–87, *88*, 94, 215n4; mobilization and, 99; moderation and, 1–2, 10, 14, 16; policy alignment and, 160; propaganda and, 40, 50, 83, 85, 122, 174; QAnon and, 10, 40, 136–38; regulation and, 49–50, 53, 55–61; research methodology and, 184; suspensions and, 125; Taliban and, 125–28; video and, 1, 50, 83, 125, 128, 136, 138, 151, 175, 178

ZeroNet, 70, 73, 87

For general queries, contact info@press.princeton.edu