Platform Governance: Online Networks and Political Extremism (Prospectus Draft)

Colin M. Henry

November 15, 2021

Abstract

Political extremism is a major threat to the stability of states and societies. In the past two decades, the vast majority of civil wars have included at least one actor from a violent extremist group. Extremist elements of political parties are ascendant in most modern democracies, winning surprising victories in the United States, Europe, and South America. And terrorist attacks by extremists are only growing, with high profile attacks in the US, France, New Zealand, Iraq, Egypt, Afghanistan, and Pakistan in just the past five years.

Do the structural features of Internet platforms make political movements more extremist? In this dissertation, I explore the how the political and discursive architecture of digital communication spaces lead online communities to embrace extreme political ideologies and adopt violent repertoires of political action. I argue that communities are more likely to radicalize on platforms governed by regimes that prioritize the velocity of information. Across three theoretical and empirical chapters, I will compile ethnographic and quantitative evidence that the political processes and structures of platform governance determine which communities radicalize, and which movements embrace extremism.

Reality is created out of confusion and contradiction, and if you exclude those elements, you're no longer talking about reality. You might think that — by following language and a logic that appears consistent – you're able to exclude that aspect of reality, but it will always be lying in wait for you, ready to take its revenge. … The sad fact is that language and logic cut off from reality have a far greater power than the language and logic of reality – with all that extraneous matter weighing down like a rock any actions we take.

– Haruki Murakami, from an interview with Aum Shinrikyo terrorist Ikuo Hayashi (2000)

Political extremism is a growing threat to both the domestic and international security of states. Rebel groups expressing extremist ideologies make up the majority of combatants in civil wars post-2003 (Walter 2017). Nearly every consolidated democracy on earth now contends with political parties that officially endorse extremist ideological positions (Gidron and Ziblatt 2019), and many face surging public support for violence against domestic political opponents.[1] Perhaps most troubling, extremist movements are increasingly expanding beyond national borders, using the Internet as a communication tool to build transnational organizational and membership capacity (Caiani and Kröll 2015; Perry and Scrivens 2016). Networked communication technology like the Internet and its various platforms provide many benefits to extremists: recruitment, organization and mobilization, and, frequently, concealment from state surveillance. However, networked telecommunications platforms are more than just a tool of coordination. More than past revolutions in communication technology, the interactive nature of the Internet creates an important and potent arena for the building and reproduction of extremist ideologies and collective identities (Perry and Scrivens 2016; Bowman-Grieve 2009; Futrell and Simi 2004). The opportunity for movements to rapidly build strong collective group identities and provide spaces for new members to participate in ideological formation has produced more resilient and, at times, violent extremist movements that persist across many Internet platforms.

Of course, the idea that Internet platforms are extraordinary tools for the organization and mobilization of social movements is nothing new (see Breuer, Landman, and Farquhar

---

[1]See, for example, (Uscinski et al. 2021; Westwood et al. 2021) which show support for political violence among Americans well beyond the thresholds needed to support large-scale one-sided violence.

2015; Howard and Hussain 2011; Tufekci and Wilson 2012; Gohdes 2015). Insights earned from research on the use of Twitter, Facebook, and Google during the Arab Spring uprisings in the early 2010s is practically memetic at this point. But despite laudatory "old media" coverage from the era, empirical research on Internet organization and mobilization tells a more tepid tale. Early work identified Twitter and other mainstream social media platforms as effective means for social movements to cheaply reach many sympathetic citizens. Movement leaders also benefited from the organizational capacity of these platforms to mobilize participants in individual protest events, leverage existing ties between protesters, crowd-source information on state or counter-protest tactics, or coordinate the allocation of resources.

However, major questions remain. The Internet is a great elephant, and we know little about the political activity that occurs outside the ear we have selected for analysis. Big social media platforms, specifically Twitter and Facebook, are the primary subjects of much of the literature on political violence and communication. But social movements, and in particular extremist movements, often use a wide variety of online systems beyond traditional social media platforms to exchange ideas and build group identities. These can include "legacy" protocols like image boards, link aggregators, and simple hypertext; end-to-end encrypted messaging services; or person-to-person ("p2p") federation services. Extremist movements are, either through coordinated effort or emergent behavior, constructing networks across multiple platforms or protocols, each with their own combination of preferred media, rules, and norms. While have some understanding of how the formal and informal rules of Twitter and Facebook structure conversation, organization, and mobilization, we have little insight into how the rest of the Internet operates, either as individual platforms or as a system of multiple platforms.

How interactions take place on these platforms also warrants closer observation. More than the media of past technological revolutions, Internet communication is deeply participatory. We know from research on sexual violence, child soldiers, and other forms of

one-sided violence that coercing new recruits into costly participation builds group cohesion, strengthens collective identity, and increases the likelihood of future participation in violence. Parasocial interactions between extremist "fans" and "influencers" bears close resemblance to these processes. Rather than passive consumption, membership in social media and other platforms often requires those within the extremist movement to engage in relationship and ideological maintenance. These interactions can be rather costly: time, money, and social isolation are all costs paid by members of the extremist movement well before engaging in "offline" actions. Because these costs can be paid repeatedly and with little friction between interactions, the opportunity for movement leaders to rapidly build cohesive groups and strengthen identities is enormous.

The aim of this dissertation is to understand how the modern Internet orders extremist movements. This project argues that the structure of communication technology determines the organizational and ideological structure of successful extremist movements. In particular, it asks: does the architecture of online social networks make digital communities more extremist? And do these network structures make extremist movements more violent?

Across three chapters, I examine these questions using evidence from primarily English-speaking extremist movements from the past twenty years. In the first chapter, I build a new dataset combining existing data from social media platforms with an original digital ethnography of extremist groups across "legacy" protocols, messaging services, and p2p platforms. This chapter will help us understand the depth and breadth of the extremist problem online by constructing a "bestiary" of extremist groups and classifying platforms across a handful of conceptual dimensions. In the second chapter, I use this dataset to understand the lifecycle of extremist groups on these platforms. In this chapter, I will show that the extremist groups that persist and grow are those who integrate the organizational practices and ideological structures common to cults that allow them to effectively leverage the formal and informal rules of communication technology. In the third chapter, I examine how extremist movements maintain their group network structures across multiple platforms.

This chapter will show that while some platforms are better at blunting extremism than others, the extremist threat is the "network of networks" that persists across numerous spaces online.

Background: Online Extremists and Cults

In the spring of 1995, ten members of the religious group Aum Shinrikyo boarded trains on three lines of the Tokyo subway during the morning rush hour. Most of them were from middle-class backgrounds and held graduate degrees in fields like medicine, artificial intelligence, or robotics. All of them were carrying plastic bags of homemade liquid sarin. At a pre-planned time, each member dropped the sarin on the floor of the train car, punctured the bag with the sharpened tip of an umbrella, and exited the subway station. Their actions killed 19 and injured nearly 7,000 commuters — the deadliest non-state chemical weapons attack in history. Why did Aum become violent? Research in cultic studies has focused on the individual psychology of group members, who were socially disconnected, suffering from a variety of mental health issues, or otherwise vulnerable to recruitment by organizations promising acceptance and purpose. Other work has pointed to the unique psychological effects of "guruism," the parasocial relationship between group members and the charismatic leader. But research on terrorism and extremism in political science has spent little time on Aum or groups like it.

Conflict studies of non-state armed groups focus on groups with recognizable, rational motivations for challenging state power and engaging in political violence. NSAGs might form and pursue violent strategies to penetrate exclusionary political or economic systems, respond to repression, reclaim territory, control illegal markets, or exploit political opportunities. But extremist groups like Aum don't seem to have intelligible motivations or preferences. The Tokyo subway terrorists wanted to kill Japanese citizens, they claimed, to kick-start a Third World War, usher in a nuclear apocalypse, and receive spiritual power from the group's

leader. Extremist groups express fears of group threat that are disconnected from political or economic reality, and express preferences that do not relate to the material threats members actually face or reasonable political outcomes.

In some small part, all extremist groups share these characteristics with cults like Aum Shinrikyo. In this background section, I introduce a definition of extremism that incorporates elements of social identity theory and ideology. I explain how cults — alternately known as "high control," "high demand," or "new religious" groups — fit within a subset of this definition. Finally, I give brief overview of the main theoretical argument of the dissertation: that some online platforms reward extremist groups who possess or adopt the characteristics of cults, and these extremist groups grow more rapidly and become more violent than others.

What is extremism?

Extremism is defined by two main concepts (Berger 2018). First, identification with the in-group is inseparable from active hostility to the out-group. Typical political competition within a set of recognized rules or norms is generally not extremist, as in-group success can occur without hostile actions or violence against the out-group. The need for hostility in this case is unconditional. Second, extremist movements view the threat of the out-group as a constant crisis. The out-group could simply be moving to disadvantage in-group members, or corrupting in-group ideology or practices. Some extremist movements imbue this crisis narrative with an apocalyptic threat; the world outside the movement is in rapid, existential decline. Only membership in the extremist movement can ensure survival.

Until relatively recently, work in political science defined "extremism" as simply views significantly outside the mainstream (Wintrobe 2006). In this early conceptual structure, what constituted extremist views or who counted as an "extremist" meant identifying a central tendency on a spectrum of political outcomes and then deciding on a threshold distance from this centre that is sufficiently "extreme." This formalization of extremism had the advantage of readily admitting itself to integration into existing rational actor models

of terrorism and political violence. Unfortunately, it also had the disadvantage of conceptually identifying almost nothing of empirical value. Genocidal or monocidal preferences are captured alongside policy preferences outside the status quo but nonetheless within the existing regime of political competition (e.g., preferring a much higher or much lower tariff). Yet few of us would consider opposition to or support for a tariff to be an extremist position. Worse still, research into radical Islamic groups in the aftermath of the September 11, 2001 terrorist attacks in the United States led to a widespread convergence of "extremism" with related terms like "terrorism," "radicalization," and in some unfortunate cases even "political violence." This phenomenon was mostly sociological. "Anti-terrorism" research on "countering violent extremism" (CVE) was a gold rush for government-funded research in US and European institutions, and produced a cottage industry of "terrorism-explainers" who mostly said things intended to be useful and flattering to US intelligence and security agencies (Wilson 2021).

One difficulty in nailing down a specific definition of extremism is that the term is simultaneously used at three different levels of analysis: individual preferences for extreme political outcomes, organizations who advocate for extreme policy positions or tactics, and movements with extreme goals. In this dissertation, I focus on movements, rather than individual psychology (as an outcome) or organized groups (like groups within the American militia movement or under the umbrella of Islamic extremism). Movements are amorphous and loosely affiliated; "membership" in a movement does not require the kind of high-cost commitment that organized group membership requires. Extremist movements based around shared group identity encourage active participation in the ideological or myth-building process as a means of policing group boundaries, emphasizing hostility towards out-group members and information, and portraying any out-group threat as overwhelming or apocalyptic. These processes are all enhanced and strengthened in online settings where network structures encourage continuous interaction and maintenance of parasocial relationships. The threat is the resilience and operation of the network, not any single individual platform,

group, or member.

Extremist movements and cults

One of the central arguments of this dissertation is that the structural features of online platforms reward extremist movements that share characteristics with cults. Cults are a specific type of extremist group: they rigorously police in-group identities. Membership in a cult almost universally requires renunciation of the rest of the world, which is viewed with suspicion, indifference or animosity, and as a threat to the purity of the internal community. Furthermore, extremist groups view the threat of the out-group as a constant crisis. The out-group could simply be moving to disadvantage in-group members, or corrupting in-group ideology or practices. Cults imbue this crisis narrative with an apocalyptic threat; the world outside the cult is in rapid, existential decline. Only membership in the cult can ensure survival. Because cults occupy this particular fringe area on the spectrum of extremism, they may be more prone to violent action. The combination of ideological structure, apocalyptic belief, and organizational structure of demand and control makes these groups vulnerable to escalation into violence against the world outside the cult.

Political cults are not a new phenomenon. The Shining Path under founder Abimael Guzmán was a "millenarian" insurgency that sought to annihilate Peruvian society entirely. Sendero Luminoso military schools trained soldiers in self-critical "moral purification" practices intended to bind them to Guzmán personally. The Tamil Tigers' practice of forbidding all sexual activity among members – a canonical example of how rebel groups can prevent rape during civil war in the sexual violence literature – was also a classic isolationist tactic meant to break down relationships that could threaten devotion to the group. Other modern examples include Aum Shinrikyo, the Romanian fascist cult the Iron Guard, the secretive Mexican cult El Yunque, the international Rajneesh community, and the burgeoning QAnon movement.

Most of what we know about cults comes out of the small but growing cultic studies

subfield in sociology, or out of the behavioral psychology literature on abuse and religious conversion (both literatures reject the pejorative term "cult" and refer to these groups as "high demand" or "high control"; for simplicity, however, I simply use the word "cult" in this proposal). Political science has mostly ignored cults as distinct phenomena. One explanation is that some cults lack the clear political objectives of parties, protest movements, rebel groups, or other traditional political actors. Organizations like the Church of Scientology that blur the lines between cults and new religious movements engage political institutions only as a means of preserving autonomy within the state, rather than accomplishing specific policy goals. Of course, this distinction has not deterred scholarship on organized criminal groups or other actors just outside the political mainstream. Perhaps it is because the process of joining a cult confounds our rationalist explanations of recruitment and radicalization. Joining Aum, for instance, required the recruit to "renounce" the world, cutting ties with family and friends, quitting any job, and transferring all wealth to the group – including valuable real estate in the midst of a massive housing boom. In return for this great cost, they received starvation rations, sleep deprivation, and exposure to a variety of dangerous pseudo-scientific "treatments." Even the ideological benefit was limited; only a small part of the Aum membership was prophesied to survive the apocalypse and live in paradise with the guru.

Digital communities

In the previous sections, I described how extremist movements are built around closely-held identities and fueled by interactions that demonize, often violently, out-groups and reinforce anxieties about in-group boundaries. I also explored how cults are a type of extremist group, characterized by apocalyptic threat from the out-group, a constant crisis of in-group cohesion, and the supplanting of all other identities by the cultic identity. How do internet communication technologies fit in? Plainly put, some digital platforms — specifically, those that feature recommendation algorithms — reward communities that share characteristics

with cults. In these spaces, movements that share information designed to outrage in-group members, enhance the threat of the out-group, and rigorously police in-group boundaries get more attention and hold more power. By virtue of this structural feature of personalized social media, successful movements are those that radicalize faster and adopt the organizational strategies of cults to prioritize identification with the extremist movement. Extremism that emerges from these digital spaces is thus more prone to violent action and may produce more dangerous extremist organizations, like militias or other armed actors.

Much of the work on the effects of social media on beliefs and violence focuses on "silo" effects — repeated exposure to extremist content from people that think like you do. However, we know that exposure alone is not enough. The mechanism is insufficient in a few ways. First, and most importantly, it fails to contend with the primary innovation in digital media: social interaction and parasocial relationships. Unlike passive consumption of extremist content — through, say, movies, television shows, magazines, and so forth — digital spaces allow group members to immediately and publicly reveal preferences and ideological positions in response to elites and each other. It also gives low-status members the perception that they have personal relationships with high-status members. We know from research on groups in other substantive areas that these kinds of connections can be powerful drivers of group and identity formation. For example, individuals with unclear status in uncertain situations — like child soldiers (Cohen 2013), foreign fighters (Moore 2019), or even college students (Sanday 1992) — seek out costly modes of participation to demonstrate loyalty to the group. Members of digital communities behave similarly. In these contexts, identity and status are even *more* socially contingent, as self-presentation is constructed through profiles and interactions without visible or non-fungible characteristics. In other words, participants can easily hide elements of their identities and select ones to present.

Online communities make up the larger social movement, and are the primary focus of the dissertation. In the following three papers, I explore how online communities may radicalize under the rules and norms of some types of platform governance. In particular,

I argue that regimes that incentivize communities to adopt the organizational features and ideological commitments of cults are more likely to foster extremism. I also explore how interdependence between online communities can lead to radicalization even if the political institutions of the platform are effective at deterring the production and spread of extremist content.

Paper 1: Mapping the Networked Extremist Space

In the preliminary data collection project of the dissertation, I explore the universe of extremist movements and the digital spaces that they inhabit. Much has been written about mainstream social media platforms like Twitter and Facebook. However, little work has been done to characterize and categorize the wide variety of social media spaces beyond these two companies — to say nothing of online communication platforms that are *not* social media but are nonetheless widely used. Similarly, extremist researchers have focused mainly on the most visible extremist groups. More research has explored Islamic extremists in the Middle East (e.g., ISIS or al-Qaeda) than the rest of the population of extremist movements combined. While this has started to change with renewed attention to United States-based white nationalist and armed militia movements following the January 6, 2021 insurrection, a full accounting of the wide variety of extremist movements that persist online across the globe has not been done. This research gap has generated a kind of "iceberg" problem, where a small proportion of the groups and platforms has come to represent an unobserved population of extremists and communication spaces. This iceberg has wide-ranging policy implications for the states and societies attempting to navigate digital waters and contend with political extremists.

In this paper, I propose a set of characteristics to classify extremist movements by their organizational and communication strategies, membership structures, and ideology. I also propose a set of characteristics to categorize online platforms by how they generate engage-

ment, engage in content moderation, and structure conversation through formal and informal rules. This strategy of category identification will uncover the theoretical structure and empirical characteristics of online *extremist communities:* individuals loosely affiliated by their belief in extremist ideologies who inhabit and interact within a digital space, subject to a specific set of formal and informal norms. These rules order not only how information is exchanged and community members interact, but also the shared ideologies and identities that community members possess. They may be imposed by the architecture of the digital space: by the friction introduced when presenting new information or responding to others (e.g., threaded responses instead of "de-linked" or "unchained" conversations), or by the visible rules of hierarchy that order information (e.g., "retweets" or "upvotes," direct moderation action, and so on). Informal norms may also arise from the production of extremist ideologies, such as identity-signaling *shibboleths* (e.g., the Soviet flag emoji among "tankies"[2] or the "patriot" moniker among Qanon adherents), common memetic language or slogans, or forbidden topics. This data collection project will produce a hybrid dataset that fuses "big data" collection of interactions within online communities with contextual, ethnographic information about the structure of online platforms and the norms that govern the individuals who use them. In the following sections, I describe in detail the data collection method, tools, and data product.

Digital Ethnography as Method and Approach

Digital ethnography[3] refers to a wide variety of research methods that marry online data with the rich contextualization common to 'traditional' ethnography. Ethnographic studies are typically conducted within the environments and contexts of research subjects in order to generate understanding about communities, behaviors, and social or political meanings

---

[2]"Tankies" refers to an online community of extremists who support the use of violence ("send in the tanks") to create Soviet-, Chinese-, or North Korean-style Communist states.

[3]A brief survey of the literature reveals a number of labels for the same kind of approach: virtual ethnography (Hine 2000), cyberethnography (Robinson and Schulz 2009), or netnography (Kozinets 2010), for example. These are all terrible names, and I'll simply be using 'digital ethnography,' following Murphy (2008).

(Geertz 2008). Most often, researchers do fieldwork — engaging in participatory observation, where scholars follow, observe, and participate in the everyday life of the community they study. The researcher's role can range from outside observation to full participation (Madden 2017). But as more important aspects of political organization and behavior have moved into digital spaces, scholars have embraced new, innovative ways of doing "fieldwork."

Howard's network ethnographic study of "hypermedia"[4] is a blueprint for this kind of innovation in digital ethnography (2002). The author adapts the meaning of "field sites" — typically a physical or institutional space researchers visit for observation or participation — to include the important nodes in a digital social network. Sites may still be physical locations, like data centers or the offices of a technology company. More likely, they are digital spaces where communities convene, like forums, collections of websites, or even social media platforms more broadly. Compared to selecting a single physical field site, such as a village or government agency, choosing a handful of important nodes allows the researcher to keep the data of interest constant (the behaviors of and interactions between community members) while thinking of the community as having many constituent parts (see Howard 2002, 561).

In this data collection project, I conceptualize the important nodes of online extremist movements at two levels: first, the social networking service (hereafter SNS); and, second, the extremist community within the SNS. The number and distinctness of communities varies by SNS. Twitter, for example, contains a multitude of rather fuzzy communities that may be uncovered by understanding follower and conversation networks. Reddit, on the other hand, has distinct communities, complete with names, designated elites with governing power, and formal rules. The first task of the project will be the identification and classification of SNS platforms. While some digital spaces are extensively studied, others remain hidden from the gaze of extremist researchers. Many, for instance, acknowledge that Twitter and Facebook groups were used extensively for the planning and mobilization of the January 6, 2021 insur-

---

[4]"Hypermedia" is an unnessecarily techno-utopic term for the transition of print and television media to the internet.

rection in Washington, DC. Less well-known is the obscure forum Kiwi Farms, a haven for a wide variety of extremist groups and the primary publisher of the livestream and manifesto of the perpetrator of the 2019 Christchurch terrorist attack in New Zealand. Many of more violent strains of extremism originate and propogate from these less "mainstream" SNS platforms. Without observation of what these spaces look like and how they are constructed, we risk missing the roots of modern online extremism.

This task will proceed in two phases. First, discovery: adapted from Ober and Wildman (2015), a digital platform is an SNS if:

- it is an interactive Internet-based application.
- it is composed of user-generated content (UGC).
- users create and maintain service-specific profiles.
- the service is primarily used to connect users to other individuals.

Note that this definition is distinct from the broader *platform,* which can include any online, data-driven service. A shopping website, for example, is a platform, but not necessarily an SNS. The focus on specific *services* and their function in this case is deliberate. The primary service of the platform should be connection between users through engagement with UGC. This definition should be broad enough to include obvious candidates like Facebook as well as decentralized or federated services like Telegram or Mastodon. It should also exclude minor services like the comment functions on news media websites or store-and-forward messaging protocols like email.

Theorizing platform governance

Now, a close reader might find these definitions rather oblique. Too much of the discussion about how to classify and understand distinctions between online spaces is bound up in the political and economic power of the corporations that operate them for-profit. The term "platform," although adopted by many critical researchers of online politics, is itself a term

of art largely coined by these private companies. "Platform" is sophisticated enough to distinguish more complex services from simple websites, but is anodyne enough to facilitate arguments against holding tech companies responsible for content, user behavior, and, more broadly, governance. Although ostensibly it refers to a flexible server codebase on which new features can be tested and deployed (Bogost and Montfort 2007), "platform" has been deployed strategically by corporations to neutralize legal liability as well (Gillespie 2010).

Similarly, the idea of "platform governance" has been used both by scholars interested in understanding how companies in control of platforms are political actors, and by the companies themselves — again typically as a marketing term or to avoid legal exposure (Gillespie 2018). Most relevant to this dissertation is how the term "platform governance" is used to understand how systems of content moderation, interfaces, algorithms, and other socio-technical infracstructure regimes form a recognizable system of political governance (Grimmelmann 2015). Althought it is true that platform companies are themselves subject to the exercise of political power and regulation that influence their own conduct of governance (Klonick 2017), for this dissertation I only focus on the user-facing structures of governance that platforms are responsible for.

In platform studies, this idea of platform governance proceeds mostly by analogy: tech companies that control platforms are similar to states and engage in state-like behaviors that structure social networks and interaction (Van Dijck 2013), extend existing political and economic systems into the digital realm (Nieborg and Poell 2018), reproduce and encode racism and other forms of bias (Noble 2018), or consider "code as law" (Lessig 2000). This is thoughtful work that seriously engages the idea that platforms are power-neutral tech. However, much of this research interacts with political science work on states and state power — the other side of the platform governance analogy — in uneven and puzzling ways. The central focus of platform studies seems to be how tech is used to reinforce existing offline power structures, which is important work. Platforms and tech companies are active agents in an ongoing political process. But when platforms are considered governments in

and of themselves, these analogies to state-like institutions break down. States monopolize violence; what are platform companies monopolizing? State-like entities have control over the distribution of political goods; what are the goods distributed by platforms? And what would a "state-building" process for digital platforms look like?

To be sure, the answers to these questions are not readily available in the political science literature. When we think of state-like activity by private actors, we mostly focus on the material consequences of competition over violence or contestation over accessing and providing public goods. The state-building literature provides a few notable explanations for how states emerge from competing over property and taxation, like the Olson's stationary bandit (1993), the violent entrepreneur (Kurrild-Klitgaard and Svendsen 2003), or dispute mediation as a protection racket (Skarbek 2011). Broader histories of state consolidation in pre-modern Europe emphasize how different combinations of proto-state institutions dealt with competition from emerging military and economic groups (see Tilly 1992; Moore 1993) . Or consider the classic Barkey study of state formation in the Ottoman empire (1994). Faced with many of the same problems of state formation in European development, the central Ottoman state co-opted private violent specialists to consolidate state power, rather than eliminating them through warfare. In each of these examples, the state is forged from and exists in relation to competitors. Without the threat of external violence or competition for legitimacy, the state lacks some subject over which to demonstrate sovereignty. State power and sovereignty command an asymmetry; they produce a hierarchical relationship between state institutions, subjects, and competitors for state power. Within this hierarchy, the state commands the use of violence: who can be killed and who can (legitimately) do the killing.

Power and sovereignty need to be redefined for the governance of digital spaces. Platform companies are sometimes faced with physical threats — see, for example, the 2018 YouTube shooting (Antony and Thomas 2010), the seizure of servers hosting the "dark web" marketplace The Silk Road (Mann and Warren 2018), or the expropriation of LiveJournal's servers from the United States to Russia (Koltsova and Koltcov 2013) — but these instances are

conceptually distinct from matters of internal platform governance[5]. Rather, sovereignty on the digital platform is power over information. We can define it acoustically: a platform exercises state-like power when they can command a monopoly on *absolute quiet* from subjects and challengers within the digital space. Platform rules and policies can be defined with reference to this theory of power. Users of SNS platforms want their information to reach the intended audience. In other words, they prefer access to the public goods the platform provides, goods like interaction, reaction, and discursive engagement. Access to and distribution of these goods is a political process, structured by the laws, rules, and norms of the political institutions that comprise the platform ownership. Power ensures that information flows unsilenced in the intended direction.

From this framework of the sovereignty of silence[6], we can understand extremists in digital communities as actors in relation to the state-like power of platform governance. Online extremists interact with platform rules and norms strategically, pursuing policy preferences both internal and external to the platform. Some platforms compete with and eradicate extremist communities who challenge the platform's monopoly on information; others, like the seventeenth century Ottomans, co-opt would-be challengers as an extension of their power over information. The pathways platforms proceed down is largely the function of how their political institutions are arranged. How democratic are the processes of governance? How exclusionary is access to the political process or the distribution of goods? What communities on the platform align their interests with those of the platform "state"? These dimensions provide a starting point for theorizing a typology of platform regime type and understanding formal and informal rules as more than just technical decisions or engineering outcomes.

---

[5]Although they are not unrelated: Nasim Najafi Aghdam, the YouTube shooter, was primarily motivated by the platform's exclusionary governance.

[6]The philosopher Byung-Chul Han amusingly coined a similar framework the "sovereignty of shitstorms" (2017), highlighting how social media virality is a challenge to the power of platform governance. "Shitstorms" threaten to flatten the heirarchy of state-like power, overwhelming the distribution of attention and engagement goods.

Classification and Coding Ontologies

The second phase is classification. One of the central arguments of the dissertation is that the structural features of a platform create incentives for communities to radicalize, adopt more extreme ideologies, and develop violent repertoires of action. Thus, designing a classification scheme that identifies variation among these features is crucial. Unfortunately, no real systematic accounting for the varieties of internal governance on SNS platforms exists. Instead, much of the focus is on the self-governance, external governance, or co-governance interactions between platform companies, states, and international bodies (Gorwa 2019). Research on internal governance, such as it has been conceptualized, is variously concerned with content moderation, algorithmic bias, privacy policies, and so on. But a broader typology that tries to understand how these various pieces of discourse architecture hang together has not emerged. The aim of the classification portion of the data project will be to construct this typology. Sample classification and coding ontologies are a work in progress, and will be the result of an iterative process involving undergraduate research assistants in the ROCCA Lab.

Data product

Big data collection projects on online platforms is largely quantitative in nature, focusing on text and digital trace data. In this paper, I propose a big data-augmented ethnography, that combines large quantities of text information about online interactions with fieldwork on the community context of a wide selection of platforms. Two data collection processes will happen in parallel: first, digital archival work conducted by ROCCA Lab research assistants using web archiving tools like PANDORA Digital Archiving System (PANDAS), Web Curator Tool (WCT), NetarchiveSuite, or the Internet Memory Foundation's Archive-It. These archival tools will allow researchers on the ROCCA team to gather publicly available text, image, and (occasionally) video data posted by users of SNS platforms. These data will be available as relational databases at the individual unit of UGC (e.g., variously, "tweet,"

"post," "comment," and so on) from which discursive social networks can be reconstructed. Second, I will engage in ethnographic fieldwork in a selection of digital communities on SNS platforms. This will largely consist of collecting and interpreting community identities, interactions, norms, and meanings through field notes. As work on this parallel track intensifies, it will also involve interviewing key members or former members of platform companies and user communities. This ethnographic work will provide the foundations for constructing a scored typology of platform regime type similar to data products on state regime types like Polity IV or V-DEM.

Paper 2: Participatory networked extremism

Which online political communities radicalize? In this paper, I argue that the structure features of social network services (SNS) determine which communities radicalize and, in turn, which movements adopt extremist ideologies and violent repertoires of action. SNS platforms use a wide variety of mechanisms and rules to "govern" their users. Some forms of these governance structures incentivize communities to adopt the characteristics of extremists, specifically cult-like groups. When platforms shape conversation and interaction through some combinations of moderation and governance processes, community identity coalesces around increasingly negative views and behaviors of non-members. This kind of discourse architecture prioritizes engagement with a handful elite users (or a single user) who gains a kind of guru-like status. Finally, it elevates identification with the online community as the central political identity of extremist members.

Using data gathered from the digital ethnography project, I show how variation in interface and application design, formal rules, and informal norms between SNS platforms — broadly thought of as "platform governance" — leads to the radicalization and extremism of some online communities. I use a networks-based approach, identifying relationships between nodes (users) connected by edges (SNS features like "follows", "friends", "retweets",

and so on). With this approach, I will show how interconnectedness between users and their willingness to engage in costly participation in the project of building the community within a particular structural context is a leading indicator of which digital communities radicalize, and therefore which movements become extremist.

Defining & Explaining Radicalization

Radicalization into extremism is the process of embracing increasingly negative views or violent behaviors towards out-groups (or both) (Berger 2018; Della Porta 2013). This process includes not just engaging in violence, but also both developing repertoires of violent action and adopting beliefs that justify violence. Unfortunately, there is little agreement on how radicalization into extremist movements occurs, or even whether this definition fits all forms of radicalization. I will focus on "ideological" radicalization, the process of adopting beliefs that justify violence and the development of repertoires of violence. New members of extremist movements adopt these beliefs and identify extremist membership as a core component of their political identities more rapidly when they are given an active, participatory role in constructing group ideology — the kind of role that the interactive discourse of SNS platforms provide.

The study of radicalization as a distinct concept grew from debates over why Western citizens carried out violent political attacks on behalf of Middle Eastern terrorist groups. The emergence of so-called "homegrown" terrorists combined with the expansion of the war on terror produced a whole field of scholarship as well as policy makers and analysts who claimed the ability to spot radicalization in progress among Muslim communities and deter extremism (Kundnani 2012). Research on social movements, in contrast to terrorism studies, sought to de-emphasize extremist violence and situate radicalization within a cycle of protest, response, and repression (see Wieviorka 2004; Oberschall 2004; Alimi, Bosi, and Demetriou 2015). Only rarely have these two approaches interacted with each other. Both perspectives share the idea that ideology — like that of radical Islam — plays a key role

in radicalization. However, the disagreement concerns what the specific characteristics of that role. 'Ideological radicalization,' for example, describes only the cognitive process of endorsing immoderate beliefs (Neumann 2003). 'Behavioral radicalization,' on the other hand, focuses on the adoption and use of violent tactics — extremism of method rather than ideological extremism (Richards 2015).

The fuzziness of prevailing definitions of radicalization reflects this lack of scholarly consensus around the phenomenon. This definition elides the main fault line: a cognitive shift to endorsing immoderate beliefs is key, but only when those extreme beliefs explicitly endorse the use of violence.

In both social movement and terrorism studies, explanations for radicalization fall broadly into two camps: structural theories and agentic theories. Research on the former in both social movement and terrorism studies perspectives tend to focus on the responses radical or extremist movements perceive in their political environment, cycles of repression from the state, and interactions with allies.[7] Weak and unstable regimes undergoing power transitions or overseeing social transformations are vulnerable to the spread of radical movements (Kriesi 1995), particularly if states pursue strategies of political exclusion (Della Porta 2006). Although purporting to study a specific type of political violence, much of this work does not stray far from the structural literature on civil war. For example, some civil war research suggests that intermediate regimes — neither harsh autocracies nor liberal democracies — are the most likely to experience civil war (Francisco 1995, @hegre2001toward). Semi-democracies undergoing a transition from autocracy to consolidated democracy face weakening institutions and increasing the risk of civil war (Shain, Linz, and Berat 1995, @tarrow2011power). Transitional regimes may also inconsistently apply repressive tactics and grant access to governmental institutions, generating resentment that leads to civil war onset (Lichbach 1987, @moore1998repression). Do regime-level structural characteristics lead to radicalization or widespread political violence? They may, if institutions are weak,

---

[7]Foundational work on this includes Tilly (1977), McAdam (2010), and Tarrow (1989).

21

or if repression is applied inconsistently; but cross-national measures of state structure have weak explanatory power.

Early social movement studies on agency focused mainly on psychology and the psychopathologies that might drive individuals to radical and violent movements; this works has found little empirical purchase, likely because extremist groups are often able to select for certain skillsets and personalities when recruiting (Crenshaw 2010; Horgan 2008). Other theories have focused on individual motivations. Some radicalized violent actors are ideologically driven, motivated by religious or pseudo-religious doctrine (Bjørgo 2004) while others may be seeking material resources as a substitute for other types of economic activity (Weinstein 2006).

Rational choice explanations for radicalization — the most prevalent in political science literature on terrorism — focus mostly on structural explanations. In this framework, individuals are reluctant to commit resources or take risks that benefit a radical group as a whole, while shouldering the costs individually (Weinstein 2006). Of course, we observe radicalized members of violent groups engaging in costly acts fairly frequently. The classic rationalist solution to this problem is some kind of coercion, either through internal norms of true believers or through group sanctions constituent of repeated interactions between members (Pape 2003; De Mesquita 2005).

However, these explanations fail to account for the relational nature of extremist groups. These frameworks treat radical beliefs — and ideology more broadly — as purely instrumental, tools used by leaders to enforce rules or coerce member commitments. Ignoring the content of beliefs or the process of changing beliefs, it is difficult for agentic or structural explanations to grapple with the social ties within groups that make inter-group sanctioning possible. Often these ties are treated as "pre-existing." Or, worse, rationalist explanations for extremism define these positions using a spatial ideological model, treating all preferences in some "corner" of a given dimension as "extreme" (Wintrobe 2006). And, of course, these theories struggle to explain political violence committed by extremists that are non-members

or merely loosely connected to extremist groups.

Participatory networked radicalization

What non-relational models of extremism and radicalization into extremist movements miss are how identities and ideologies are constituted through repeated interaction. Journalists and activists studying extremists often tell a story about radicalization on social media platforms: people are drawn into extremist movements through a slow process of conversation and research with existing members. New members often start as skeptics who question extremists and are directed to "do their own research." A cycle of discovery, discussion, and reinforcement begins that makes targets active participants in their own recruitment, building and strengthening social ties to existing members through a process of learning and myth-making. When members are given an active role in constructing the mythology that surrounds a group's ideology, they begin to see group membership as a fundamental, core component of their identity. When this important part of their political and personal identity is questioned, they are unlikely to respond by rejecting the identity. Instead, active participants perceive threats to the group — and thus opportunities to demote group membership as an identity or relinquish belief in dis-proven ideas — as much more costly. Leaders can bind group members to the group much more tightly by granting them this active, participatory role in ideological construction, often without reducing their own power.

Approaching extremist communities and movements as networks gives distinct advantages over non-relational approaches. Recent work shows that understanding interpersonal relationships and the context in which they emerge and operate helps us understand complex outcomes. For example, Dorff's work on victimization in Mexico leverages kinship networks to understand how the complexities of interpersonal relationships mediates the link between experiencing violence and engaging in political participation, filling a crucial gap between experience and behavior (' Dorff 2017). Conversely, isolation from family and friend groups — like Israeli checkpoints that weaken ties between Palestinian communities (Gade 2020) —

23

can induce shifts in preferences towards violent behavior. Understanding how social cohesion emerges from these kinds of networks can help us understand how information travels in a system where transmission is a function of trust, and trust is constitutive of group identity.

Models of information transmission within ethnic groups show that the structure of a community network matters for how far and how quickly news spreads (Larson 2017; Arias et al. 2019). There are a few key insights from this model. First, any hesitation in sharing information with another community member can hamper the speed of information traversing the network, despite the density of the network. In practice, this means that homogeneous groups with fewer social relationships can pass along a message faster than heterogeneous groups with many social relationships. Second, network members must be fairly adept at identifying in- and out-group members for the purposes of passing along information. Information from in-group members may be distrusted by out-group members, and in-group members have incentives to misrepresent information to out-group members (Minozzi et al. 2020). And they must be able to act in relative concert (which may be an emergent property of the group rather than an intentional or strategic action). Often, this is facilitated by elite or expert appeals to in-group members to cooperate (Chang and Peisakhin 2019). The one way the first mechanism works as described is if members of one ethnic group can readily recognize the identity of another ethnic group, and if the meaning of that identity is relatively stable across all the members of the former group. Otherwise, if a small number of members of the informed group pass the information along to the uninformed group, then the density of the sub-group network would speed information transmission as expected and erase the benefits we observe in sparse, homogeneous networks.

The poster's gamble[8]

Understanding how network ties that cross in- and out-group identities and how trust modifies the acceptance of new information, especially in the context of misinformation, is an

---

[8]"The poster's gamble" is a term coined in a Tweet by Slow Fawns frontman and *Reply All* podcaster Alex Goldman (2021).

open research question (Larson 2021). Now, extremist ideologies and the interactions that sustain them are not necessarily misinformation (although extremists nearly always share ideological commitments with conspiracy theories, cultish prophesies, and urban legends). But creating, reproducing, and sharing extremist ideology is a high-trust interaction. Ideology requires interaction and informational exchange; without transmission and narrative, there is no extremist group. Sustaining a movement requires continuous participation between in-group members to build group identity and constructing ideological commitments.

"Participation" in violent or extremist groups can be costly. Rebel leaders who abduct children often force them to commit acts of violence to build their allegiance to the group and sever ties to their previous lives (Cohen 2017). Sexual violence by new rebel fighters, and in particular gang-rape, are very costly types of "participation" that reify group identity and strengthen group cohesion (Cohen, Green, and Wood 2013). Violent participatory acts like this are undertaken in view of other group members to signal resolve to the cause or willingness to sacrifice on behalf of the group (Cohen 2013). Perpetrators may receive praise from leaders or comrades, or become ensnared in a sense of collective shame that increases the personal cost of returning to non-violence life outside of the group.

However, participation can be costly without being violent. New entrants into extremist movements, like cults, are often expected to engage in low-level volunteer work, learn "educational" material, or produce new work that contributes to the ideological base of the organization (Lalich 2004). Members or adherents are tasked with seemingly meaningless work that is imbued with a sense of importance. By completing these participatory tasks and seeing their results praised by other members, new adherents obtain a growing sense of shared identity (Jasper 2011). If the movement is threatened or the ideology of the group rebutted, the threat is not just to the movement broadly, but to the individual specifically. Participatory radicalization is able to individualize group threat, which provokes a stronger response in the group member.

Social cohesion within groups is built through costly demonstrations of commitment to

the group's cause and fellow members. Nearly all work on how intragroup ties are built in violent groups has focused on physical acts — shared combat experiences or public acts of sexual violence, for example. But the construction of extremist beliefs and ideologies, and the public demonstration of beliefs changing, are also costly signals that can build social relationships between group members. Participation in extremist discourse has the dual effect of both building social ties between radicalization targets and existing group members, and constructing the ideological content of the extremist group. Expressions of belief are thus both dramaturgical and sincere political commitments.

How do these expressions happen within digital spaces? In the first paper of the proposal, I discuss how digital platforms operate a kind of "sovereignty of silence" through the political structures and processes that order communication between users. In this framework, individuals and groups act strategically to achieve the greatest reach for user-generated content (UGC). In other words, members of digital communities who post new information or participate in conversations have the same general goal: they want others to view what they have written or created, and respond in some way. This may be purely competitive, as many have theorized (Hwang 2020; Burrell and Fourcade 2020); but it may also be cooperative, with multiple users "boosting" an idea to reach a greater share of the platform audience. The interactions that build and sustain extremist ideologies operate similarly. Members of radicalizing communities face a kind of "poster's gamble:" participating in extremist discourse is costly, and the payoff depends on the audience reward system devised by the owners of the digital platform. On platforms that prioritize information velocity, UGC that is *more extremist* is more likely to provide an attention payoff to the user, solidifying their status within the group. Members of radicalizing communities are incentivized to produce more extreme content more often, contributing to the reification of the extremist group identity and the radicalization of the community more broadly.

Initially, I argue that platforms that prioritize information velocity will have a collection of common characteristics. These spaces will have governance regimes that are more

obfuscated, arbitrary, less democratic, and with fewer avenues for redress once subjected to platform power. As the ethnographic and textual data collection processes in the first paper of the dissertation proceed, I expect to solidify the characteristics of distinct platform regime types. I also expect that political communities on platforms with these common characteristics will radicalize faster over time. Specifically, these communities will embrace extremist ideologies that resemble cultic belief systems: featuring apocalyptic threat from out-groups, violent and often monocidal repertoires of action, and more vigorous policing of group boundaries.

Empirical strategy

To begin exploring these theoretical expectations, I turn to the loosely organized extremist "Qanon" movement in the United States. Adherents to this movement believe that violence is necessary to combat an international cabal of sex traffickers that control the United States government. The movement has no physical organizing space; the vast majority of adherents are radicalized through online communities. While these users do advocate for the violent overthrow of democratically elected leaders, they are in some respects less radical than more established long-time right-wing terrorist groups. This latter category consists mostly of white nationalist, sovereign citizen, neo-Nazi, and other anti-state groups that have long cultivated criminal networks to build resources and train recruits to commit violent acts of terrorism. Qanon adherents, in contrast, are mostly newcomers to radical ideologies, often lacking training in violent or insurgent activities, and holding views that, while abhorrent, are not explicitly white nationalist, anti-Semitic, or anti-state. Qanon adherents are also distinguished from more organized and violent fascists in that they advocate for a third party — the US military — to engage in violence on their behalf, rather than plan for their own acts of violence. Thus, they represent an important opportunity to study the process at work between introduction to and belief in extremist ideology.

The movement is spread across a wide variety of platforms and has emerged in many

unexpected communities. To study the effects of platform structure on the radicalization of digital spaces, I propose a mixed-methods research design. First, using the dataset gathered in the first paper of the dissertation, I will select a representative sample of platforms from across the typology of platform governance regime types. For each of these platforms, I will build a series of dependent variables measuring different aspects of extremism from a combination of text and relational data gathered during the archival stage of the first paper, augmented by datasets already compiled by the Network Contagion Research Institute's Social Media Analysis Toolkit (SMAT) and the journalists at Distributed Denial of Secrets (DDoSecrets). At this time, my current collection of text, profile, and relational data includes samples from the Twitter, Gab, Parler, 4chan, 8kun, Telegram, Gettr platforms from 2016 to 2020.

This data collection approach faces two main obstacles. First, detecting and measuring extremism is not a solved problem. The biggest tech platforms employ thousands of people and spend millions of dollars trying to detect extremist content at scale, and have largely failed. I propose a series of measures for extremism that might otherwise be unworkable in the real-time detection systems used by online platforms. The first is concerned with *adoption of extreme ideologies:* how often and how central to discussion are topics related to threat from the out-group, violence towards out-group members as matters of policy, reverence of a political figure, and rigorous boundary policing. The presence of these topics in single posts and their topical centrality in the network of a particular discussion can be obtained using topic modeling and network text analysis. The second focuses on increasingly *strong acquired group identity*: the use of linguistic *shibboleths,* self-identification in SNS platform profiles, and the repression of other identifiers. Again, these elements can be detected using simple text analysis techniques. Using these two elements, users and their ego-nets can be scored based on their use of extremist ideological rhetoric and commitment to shared group identity in the extremist movement.

The second obstacle is that representative sampling across multiple SNS platforms is

difficult. Frankly, sampling within a single platform is a contested concept, and debate continues on how to effectively sample even a mainstream platform like Twitter (see Kim et al. 2018 for an overview). To be clear, the empirical strategy in this paper only seeks to compare users sampled from communities on multiple platforms, rather than building a dataset composed of users and UGC from many different SNS platforms. Still, traditional techniques of purposive and probability sampling can be polluted by the algorithmic ranking and provision systems of SNS platforms themselves. I propose using a mixed-methods sampling technique first suggested by Teddie and Yu (2007). I will construct a sample of users and UGC from communities stratified by the size of community membership relative to the size of the platform userbase, activity rate, and topical "closeness" to existing types of extremism.

Using these data, I will then build models of radicalization at several different levels of analysis. The first will look at the effects of platform-wide governance regimes on two measures of extremism, *adoption of extreme ideologies* and *strong acquired group identity.* This model will examine differences between platforms from a variety of governance regimes across the typology constructed in the first paper of the dissertation. The second will examine the effects of changes in governance within a particular platform across time on the same two measures. Some platforms, like Reddit, change their processes and structures of governance frequently and publicly. I will leverage this variation to understand differences in radicalization of some communities within the platform.

Next steps

All of the empirical strategy and some of the key theoretical concepts in this paper depend on the progress of the first paper in the dissertation. As the ROCCA team and I investigate, classify, and archive digital spaces, I will begin updating the conceptual definitions and understanding of the mechanisms at work within extremist communities. The ROCCA team's work will commence during the Spring semester of 2022, but preparation for the team

will begin well before then. The first step is to produce a first iteration of the classification ontology that we will use to process SNS platforms. In parallel, I will begin setting up the infrastructure necessary to receive, store, and access large quantities of text and relational data from a variety of sources. As we begin to get a grasp of the landscape of platforms and extremist movements out there, I will be able to start constructing a more specific mixed-methods sampling strategy. Assuming satisfactory progress with the ROCCA team, I expect to complete the empirical analysis for this paper by the end of the 2022 Spring semester.

Paper 3: Multi-platform Extremism

In the first empirical paper, I explore how differences in the structure of SNS platforms contribute to the political radicalization of online communities. Some platform structures incentivize communities to adopt cult-like organizational and ideological characteristics, which encourages members to embrace violent rhetoric and repertoires of action. But communities and platforms do not function independent of each other. In fact, one of the unique features about this generation of communication technology is the degree of network dependency between digital spaces. Information — as well as identities, ideologies, and norms — travel across connections between communities on different SNS platforms. Little research has been done on the relationships and interdependence between platforms and extremist communities on platforms.

And yet we observe these important dynamics in a variety of research arenas. In civil conflict, for instance, a hyperfocus on dyadic interactions between the state and one or two prominent rebel groups obscures the empirical fact that a third of all wars involve multiple interdependent group interactions (Dorff, Gallop, and Minhas 2020). When new actors enter or exit the war, it can shift the structure of the conflict network to induce other actors to escalate fighting or seek peaceful agreements (Gade et al. 2019). Within rebel groups or wider anti-state movements, the network of interdependent ties can help explain why groups

splinter or descend into infighting, as well (Gade, Hafez, and Gabbay 2019).

In the misinformation space one example of this dynamic at work is the case of the White Helmets, a volunteer humanitarian group operating in Syria, and the target of a disinformation campaign that spans multiple SNS platforms. Some of the only research on this type of networked information flow shows how specific ideas and images originated in some spaces (Youtube, Russia Today) and spread to others (Twitter), generating anti-White Helmets sentiment in communities otherwise insulated from disinformation (Wilson and Starbird 2020).

Does radicalization occur the same way? In other words, can political communities on platforms with "better" structural features become radicalized through their network ties to more extremist communities? In this paper, I will explore this research question using a mixed-methods approach, original data collected from the digital ethnography project, and data obtained by a collection of journalists working for the DDoSecrets non-profit.

Radicalization across multiple platforms

In the aftermath of the January 6th insurrection at the US Capitol, mainstream SNS platforms like Twitter and Facebook began a widespread purge of users who expressed support of the Qanon conspiracy theory or advocated for extremist violence against the state. Members of more extremist communities viewed this purged group as an important recruiting pool. Data from the SNS Gab obtained by DDOS shows that white nationalist and anti-semitic extremist groups deliberately employed a strategy of slowly introducing conspiracy theorists to extremist ideas and norms, encouraging them to conduct "research" on their own. As described above, this is an effective strategy because individuals who adopt extremist ideology or identities through a participatory process more readily integrate them into their core worldview and are less likely to revise these views or demote these identities when confronted with disconfirming or debunking information. Extremists often employ these strategies deliberately, coordinating in private or semi-private communication channels.

How might this work in theory? In the first empirical paper in the dissertation, I will explore how the features of a platform might induce a political community to radicalize. However, there remain significant unexplained factors outside the platform or the community itself that might explain the emergence of extremism. The story of radicalization in a self-contained platform focuses on the incentives members have to create, post, and reproduce increasingly extremist content. In this theory, members seek certainty about and enhancement to their status as members of the in-group. Attention and engagement, forms of public goods provided by the platform regime, can help determine the relative status of a member within the extremist group. If the platform prioritizes the velocity of information, then members can often achieve greater status by contributing increasingly extremist ideas or tactics to the broader project of constructing the group ideology.

However, platform features are not the only factor that can affect the distribution of audience reward in a particular community. External actors can and will leverage the same platform rules to alter the hierarchy of power within a digital community. Examples abound: in mid-2014, misogynistic extremists from a group of small, insular gaming communities online sparked a harassment campaign of female, LGBTQ, and BIPOC video game journalists. Hailing from platforms like 4chan and 8kun — those with moderation and governance structures that grossly intensify violent extremism — extremists used the recommendation algorithms and other features of Reddit to introduce extremist ideologies and identities to a variety of communities on the platform (Massanari 2017). Relational ties between extremist communities on 4chan and relatively moderate, apolitical communities on Reddit — namely on the basis of shared interest in video games — were used to create a permission structure to embrace extremist ideologies, practices of out-group harassment, and, eventually, violent repertoires of action.

We can begin to disentangle the mechanism here by considering the different status-value of information from different sources passing through a community's network. Typically, in-network interactions are the most valuable to in-group members. Community members want

to discuss topics related to the group identity and ideology, and they want fellow members to interact with these discussions to improve and solidify their status as fellow in-group members. Cross-network and out-network interactions, however, can increase in value within a community in certain circumstances. For communities that have already radicalized, reaching out to radicalize sympathetic just-out-group members (cross-network interactions) or to direct violence against out-group members (out-network interactions) can also produce status-seeking effects. In the latter instance, harassing and directing violent threats towards the perceived enemies of the extremist group is continuing to do the fundamental work of extremism. It is, in effect, the ideology at work. Extremists may engage in coordinated or singular campaigns of violence and harassment and then use these experiences to boost group cohesion or reify group identity. In the former, extremist groups can shift the centrality of extremist topics in otherwise moderate political communities. Often extremists can achieve this by using highly coordinated tactics to overwhelm unprepared systems of platform governance. By producing extremist content that appears to perform well within the community's audience reward structure, extremists can create the appearance that the "poster's gamble" requires in-group members to embrace more extremist ideas or tactics.

Empirical strategy

There is little guidance on how to study the behavior of network structures that span multiple SNS platforms and communities. However, the data collection project of the first paper of the dissertation should provide significant leverage on this problem. I see two possible ways to begin modeling these interactions: first, by tracking the movement of users across many platforms; and, second, tracking the reuse of links, images, and videos across platforms. In the first instance, I can use the data collection of journalists at DDoSecrets to begin building a dataset of extremists that deliberately cross platforms to engage in this kind of online activity. The datasets built by DDoSecrets on Gab and Parler, well-known SNS platforms for right-wing extremists, contain email addresses linked to account creation on

these platforms and others. There are significant ethical concerns with this practice, however. While it is common in digital activist circles to "doxx" extremists by tracking down the real, physical individual behind accounts that produce extremist content, that activity is not appropriate for an academic study. In preliminary talks with DDoSecrets, we have discussed a data sharing agreement in which privately identifying information (PII) might be used to construct a network dataset within a secure environment, and then any PII irrevocably destroyed.

In the second instance, the data collection and modeling process is focused on the repetition of information, rather than specific users. Digital trace data — in the form of hyperlinks, images, and videos — can be used to track how information moves between extremist and non-extremist communities. I could then score this information using the same criteria for topical extremist in the first empirical paper of the dissertation. Coupled with the knowledge of whether this information comes from in-, cross-, or out-network interactions, I could model the effects of community interdependence on the radicalization of non-extremist platforms.

Conclusion

In the previous sections, I described a series of papers that aim to understand how the structure of Internet platforms generates and sustains extremist movements. I make the argument that the structure of communication technology determines the organizational and ideological structure of extremist communities, and through them, the larger political movement. In particular, I argue that platforms that prioritize the velocity of information and have less responsive governance institutions push communities to radicalize.

Across three chapters, I intend to support these arguments using evidence from primarily English-speaking extremist movements from the past twenty years. The first chapter is a data collection project that combines existing data from social media platforms with a mixed-methods digital ethnography of extremist groups across a variety of SNS platforms. This

chapter will employ a team of ROCCA research assistants to identify, catalog, and classify platforms within a typology of online governance. I intend to complete the data collection project by the end of the 2022 Spring semester, but the on-going process of big data collection and ethnographic investigation will support completing the other chapters throughout the Spring.

In the second chapter, I will use this hybrid dataset to understand how political communities radicalize within a particular platform. In this paper, I argue that the communities that radicalize are those who incentivized by platform rules and norms to integrate the organizational practices and extremist ideological structures of cults. While this paper depends on the data collection project in the first paper, there is still significant work to be done on clarifying and specifying the theoretical process and mechanisms in greater detail. I will also be at work solidifying the complex sampling strategy that the within-platform models of radicalization will require.

In the third chapter, I examine how extremists maintain and expand their movement across multiple platforms. In this paper, I argue that while some platforms have governance structures that do not incentivize radicalization, the "network of networks" that persists across digital communities can nonetheless induce political communities to radicalize. This paper is mostly speculative at the present stage, and requires significant theorizing about the mechanisms as well as the modeling strategy.

Finally, I intend to present at the defense a fourth chapter that uses an experimental design to show that the rules of communication structures *alone* are enough to induce extremist beliefs among participants. Drawing from a rich tradition of work on discourse architecture, this paper will argue that we can get strong, causal leverage on the mechanisms that incentivize communities to draw strict group boundaries, perceive out-group members as destructive threats, and propose increasingly violent tactics against out-groups through an in-the-field experiment with Vanderbilt undergraduates.

Bibliography

Alimi, Eitan Y, Lorenzo Bosi, and Chares Demetriou. 2015. *The Dynamics of Radicalization: A Relational and Comparative Perspective.* Oxford University Press.

Antony, Mary Grace, and Ryan J Thomas. 2010. "'This Is Citizen Journalism at Its Finest': YouTube and the Public Sphere in the Oscar Grant Shooting Incident." *New Media & Society* 12 (8): 1280–96.

Arias, Eric, Pablo Balán, Horacio Larreguy, John Marshall, and Pablo Querubín. 2019. "Information Provision, Voter Coordination, and Electoral Accountability: Evidence from Mexican Social Networks." *American Political Science Review* 113 (2): 475–98.

Barkey, Karen. 1994. *Bandits and Bureaucrats: The Ottoman Route to State Centralization.* Cornell University Press.

Berger, John M. 2018. *Extremism.*

Bjørgo, Tore. 2004. *Root Causes of Terrorism: Myths, Reality and Ways Forward.* Routledge.

Bogost, Ian, and Nick Montfort. 2007. "New Media as Material Constraint: An Introduction to Platform Studies." In *Electronic Techtonics: Thinking at the Interface. Proceedings of the First International Hastac Conference*, 176–93.

Bowman-Grieve, Lorraine. 2009. "Exploring 'Stormfront': A Virtual Community of the Radical Right." *Studies in Conflict & Terrorism* 32 (11): 989–1007.

Breuer, Anita, Todd Landman, and Dorothea Farquhar. 2015. "Social Media and Protest Mobilization: Evidence from the Tunisian Revolution." *Democratization* 22 (4): 764–92.

Burrell, Jenna, and Marion Fourcade. 2020. "The Society of Algorithms." *Annual Review of Sociology* 47.

Caiani, Manuela, and Patricia Kröll. 2015. "The Transnationalization of the Extreme Right and the Use of the Internet." *International Journal of Comparative and Applied Criminal Justice* 39 (4): 331–51.

Chang, Han Il, and Leonid Peisakhin. 2019. "Building Cooperation Among Groups in

Conflict: An Experiment on Intersectarian Cooperation in Lebanon." *American Journal of Political Science* 63 (1): 146–62.

Cohen, Dara Kay. 2013. "Explaining Rape During Civil War: Cross-National Evidence (1980–2009)." *American Political Science Review* 107 (3): 461–77.

———. 2017. "The Ties That Bind: How Armed Groups Use Violence to Socialize Fighters." *Journal of Peace Research* 54 (5): 701–14.

Cohen, Dara Kay, Amelia Hoover Green, and Elisabeth Jean Wood. 2013. "Wartime Sexual Violence: Misconceptions, Implications, and Ways Forward." *USIP Special Report.*

Crenshaw, Martha. 2010. *Terrorism in Context.* Penn State Press.

Della Porta, Donatella. 2006. *Social Movements, Political Violence, and the State: A Comparative Analysis of Italy and Germany.* Cambridge University Press.

———. 2013. *Clandestine Political Violence.* Cambridge University Press.

De Mesquita, Ethan Bueno. 2005. "Conciliation, Counterterrorism, and Patterns of Terrorist Violence." *International Organization*, 145–76.

Dorff, Cassy. 2017. "Violence, Kinship Networks, and Political Resilience: Evidence from Mexico." *Journal of Peace Research* 54 (4): 558–73.

Dorff, Cassy, Max Gallop, and Shahryar Minhas. 2020. "Networks of Violence: Predicting Conflict in Nigeria." *The Journal of Politics* 82 (2): 476–93.

Francisco, Ronald A. 1995. "The Relationship Between Coercion and Protest: An Empirical Evaluation in Three Coercive States." *Journal of Conflict Resolution* 39 (2): 263–82.

Futrell, Robert, and Pete Simi. 2004. "Free Spaces, Collective Identity, and the Persistence of Us White Power Activism." *Social Problems* 51 (1): 16–42.

Gade, Emily Kalah. 2020. "Social Isolation and Repertoires of Resistance." *American Political Science Review* 114 (2): 309–25.

Gade, Emily Kalah, Michael Gabbay, Mohammed M Hafez, and Zane Kelly. 2019. "Networks of Cooperation: Rebel Alliances in Fragmented Civil Wars." *Journal of Conflict Resolution* 63 (9): 2071–97.

Gade, Emily Kalah, Mohammed M Hafez, and Michael Gabbay. 2019. "Fratricide in Rebel Movements: A Network Analysis of Syrian Militant Infighting." *Journal of Peace Research* 56 (3): 321–35.

Geertz, Clifford. 2008. *Thick Description: Toward an Interpretive Theory of Culture.* Routledge.

Gidron, Noam, and Daniel Ziblatt. 2019. "Center-Right Political Parties in Advanced Democracies." *Annual Review of Political Science* 22: 17–35.

Gillespie, Tarleton. 2010. "The Politics of 'Platforms'." *New Media & Society* 12 (3): 347–64.

———. 2018. *Custodians of the Internet.* Yale University Press.

Gohdes, Anita R. 2015. "Pulling the Plug: Network Disruptions and Violence in Civil Conflict." *Journal of Peace Research* 52 (3): 352–67.

Goldman, Alex. 2021. "Sometimes Tweets Are Good. Sometimes Tweets Are Bad. It's Hard to Know Until They're Out in the World. Scientists Call This "the Poster's Gamble"." *Tweet.* Twitter. https://twitter.com/AGoldmund/status/1451745843946692610?s=20.

Gorwa, Robert. 2019. "What Is Platform Governance?" *Information, Communication & Society* 22 (6): 854–71.

Grimmelmann, James. 2015. "The Virtues of Moderation." *Yale JL & Tech.* 17: 42.

Han, Byung-Chul. 2017. *In the Swarm: Digital Prospects.* Vol. 3. MIT Press.

Hegre, Håvard, Tanja Ellingsen, Scott Gates, and Nils Petter Gleditsch. 2001. "Toward a Democratic Civil Peace? Democracy, Political Change, and Civil War, 1816-1992." *American Political Science Review*, 33–48.

Hine, Christine. 2000. "The Virtual Objects of Ethnography." *Virtual Ethnography.*

Horgan, John. 2008. "From Profiles to Pathways and Roots to Routes: Perspectives from Psychology on Radicalization into Terrorism." *The ANNALS of the American Academy of Political and Social Science* 618 (1): 80–94.

Howard, Philip N. 2002. "Network Ethnography and the Hypermedia Organization: New

Media, New Organizations, New Methods." *New Media & Society* 4 (4): 550–74.

Howard, Philip N, and Muzammil M Hussain. 2011. "The Upheavals in Egypt and Tunisia: The Role of Digital Media." *Journal of Democracy* 22 (3): 35–48.

Hwang, Tim. 2020. *Subprime Attention Crisis: Advertising and the Time Bomb at the Heart of the Internet.* FSG originals.

Jasper, James M. 2011. "Emotions and Social Movements: Twenty Years of Theory and Research." *Annual Review of Sociology* 37: 285–303.

Kim, Hwalbin, S Mo Jang, Sei-Hill Kim, and Anan Wan. 2018. "Evaluating Sampling Methods for Content Analysis of Twitter Data." *Social Media+ Society* 4 (2): 2056305118772836.

Klonick, Kate. 2017. "The New Governors: The People, Rules, and Processes Governing Online Speech." *Harv. L. Rev.* 131: 1598.

Koltsova, Olessia, and Sergei Koltcov. 2013. "Mapping the Public Agenda with Topic Modeling: The Case of the Russian Livejournal." *Policy & Internet* 5 (2): 207–27.

Kozinets, Robert V. 2010. *Netnography: Doing Ethnographic Research Online.* Sage publications.

Kriesi, Hanspeter. 1995. "The Political Opportunity Structure of New Social Movements: Its Impact on Their Mobilization." *The Politics of Social Protest: Comparative Perspectives on States and Social Movements* 3.

Kundnani, Arun. 2012. "Radicalisation: The Journey of a Concept." *Race & Class* 54 (2): 3–25.

Kurrild-Klitgaard, Peter, and Gert Tinggaard Svendsen. 2003. "Rational Bandits: Plunder, Public Goods, and the Vikings." *Public Choice* 117 (3): 255–72.

Lalich, Janja. 2004. *Bounded Choice: True Believers and Charismatic Cults.* Univ of California Press.

Larson, Jennifer M. 2017. "Networks and Interethnic Cooperation." *The Journal of Politics* 79 (2): 546–59.

———. 2021. "Networks of Conflict and Cooperation." *Annual Review of Political Science* 24: 89–107.

Lessig, Lawrence. 2000. "Code Is Law." *Harvard Magazine* 1 (2000).

Lichbach, Mark Irving. 1987. "Deterrence or Escalation? The Puzzle of Aggregate Studies of Repression and Dissent." *Journal of Conflict Resolution* 31 (2): 266–97.

Madden, Raymond. 2017. *Being Ethnographic: A Guide to the Theory and Practice of Ethnography.* Sage.

Mann, Monique, and Ian Warren. 2018. "The Digital and Legal Divide: Silk Road, Transnational Online Policing and Southern Criminology." In *The Palgrave Handbook of Criminology and the Global South*, 245–60. Springer.

Massanari, Adrienne. 2017. "# Gamergate and the Fappening: How Reddit's Algorithm, Governance, and Culture Support Toxic Technocultures." *New Media & Society* 19 (3): 329–46.

McAdam, Doug. 2010. *Political Process and the Development of Black Insurgency, 1930-1970.* University of Chicago Press.

Minozzi, William, Hyunjin Song, David MJ Lazer, Michael A Neblo, and Katherine Ognyanova. 2020. "The Incidental Pundit: Who Talks Politics with Whom, and Why?" *American Journal of Political Science* 64 (1): 135–51.

Moore, Barrington. 1993. *Social Origins of Dictatorship and Democracy: Lord and Peasant in the Making of the Modern World.* Vol. 268. Beacon Press.

Moore, Pauline. 2019. "When Do Ties Bind? Foreign Fighters, Social Embeddedness, and Violence Against Civilians." *Journal of Peace Research* 56 (2): 279–94.

Moore, Will H. 1998. "Repression and Dissent: Substitution, Context, and Timing." *American Journal of Political Science*, 851–73.

Murthy, Dhiraj. 2008. "Digital Ethnography: An Examination of the Use of New Technologies for Social Research." *Sociology* 42 (5): 837–55.

Neumann, Peter R. 2003. "The Trouble with Radicalization." *International Affairs* 89

(4): 873–93.

Nieborg, David B, and Thomas Poell. 2018. "The Platformization of Cultural Production: Theorizing the Contingent Cultural Commodity." *New Media & Society* 20 (11): 4275–92.

Noble, Safiya Umoja. 2018. *Algorithms of Oppression.* New York University Press.

Obar, Jonathan A, and Steven S Wildman. 2015. "Social Media Definition and the Governance Challenge-an Introduction to the Special Issue." *Obar, JA and Wildman, S.(2015). Social Media Definition and the Governance Challenge: An Introduction to the Special Issue. Telecommunications Policy* 39 (9): 745–50.

Oberschall, Anthony. 2004. "Explaining Terrorism: The Contribution of Collective Action Theory." *Sociological Theory* 22 (1): 26–37.

Olson, Mancur. 1993. "Dictatorship, Democracy, and Development." *American Political Science Review* 87 (3): 567–76.

Pape, Robert A. 2003. "The Strategic Logic of Suicide Terrorism." *American Political Science Review*, 343–61.

Perry, Barbara, and Ryan Scrivens. 2016. "White Pride Worldwide: Constructing Global Identities Online." *The Globalisation of Hate: Internationalising Hate Crime*, 65–78.

Richards, Anthony. 2015. "From Terrorism to 'Radicalization'to 'Extremism': Counterterrorism Imperative or Loss of Focus?" *International Affairs* 91 (2): 371–80.

Robinson, Laura, and Jeremy Schulz. 2009. "New Avenues for Sociological Inquiry: Evolving Forms of Ethnographic Practice." *Sociology* 43 (4): 685–98.

Sanday, Peggy Reeves. 1992. *Fraternity Gang Rape: Sex, Brotherhood, and Privilege on Campus.* NYU Press.

Shain, Yossi, Juan J Linz, and Lynn Berat. 1995. *Between States: Interim Governments in Democratic Transitions.* Cambridge University Press.

Skarbek, David. 2011. "Governance and Prison Gangs." *American Political Science Review* 105 (4): 702–16.

Tarrow, Sidney G. 1989. *Democracy and Disorder: Protest and Politics in Italy, 1965-1975.* Oxford University Press, USA.

———. 2011. *Power in Movement: Social Movements and Contentious Politics.* Cambridge University Press.

Teddie, C, and F Yu. 2007. "Mixed Method of Sampling with Examples." *Journal of Mixed Methods Research. Sage Publications. US, A.*

Tilly, Charles. 1977. "From Mobilization to Revolution."

———. 1992. *Coercion, Capital, and European States, Ad 990-1992.* Wiley-Blackwell.

Tufekci, Zeynep, and Christopher Wilson. 2012. "Social Media and the Decision to Participate in Political Protest: Observations from Tahrir Square." *Journal of Communication* 62 (2): 363–79.

Uscinski, Joseph E, Adam M Enders, Michelle I Seelig, Casey A Klofstad, John R Funchion, Caleb Everett, Stefan Wuchty, Kamal Premaratne, and Manohar N Murthi. 2021. "American Politics in Two Dimensions: Partisan and Ideological Identities Versus Anti-Establishment Orientations." *American Journal of Political Science.*

Van Dijck, José. 2013. *The Culture of Connectivity: A Critical History of Social Media.* Oxford University Press.

Walter, Barbara F. 2017. "The New New Civil Wars." *Annual Review of Political Science* 20: 469–86.

Weinstein, Jeremy M. 2006. *Inside Rebellion: The Politics of Insurgent Violence.* Cambridge University Press.

Westwood, Sean, Matthew Tyler, Clayton Nall, and others. 2021. "Political Violence."

Wieviorka, Michel. 2004. *The Making of Terrorism.* University of Chicago Press.

Wilson, Lydia. 2021. "Gone to Waste: The 'Cve' Industry After 9/11." *New Lines*, 1–1. https://newlinesmag.com/argument/understanding-the-lure-of-islamism-is-more-complex-than-the-experts-would-have-you-believe/.

Wilson, Tom, and Kate Starbird. 2020. "Cross-Platform Disinformation Campaigns:

Lessons Learned and Next Steps." *Harvard Kennedy School Misinformation Review* 1 (1).

Wintrobe, Ronald. 2006. *Rational Extremism: The Political Economy of Radicalism.* Cambridge University Press.