

©Copyright 2024

Charles Kiene

Identity, Legitimacy, and Voice: Understanding Rule Adoption, Compliance, and Evolution in Online Communities

Charles Kiene

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2024

Reading Committee:

Benjamin Mako Hill, Chair

Kirsten Foot

Aaron D Shaw

Mark Zachry

Program Authorized to Offer Degree:
Department of Communication

University of Washington

Abstract

Identity, Legitimacy, and Voice: Understanding Rule Adoption, Compliance, and
Evolution in Online Communities

Charles Kiene

Chair of the Supervisory Committee:
Benjamin Mako Hill
Department of Communication

Online communities have become deeply entangled with 21st century social life, supporting cultural discourse and the production of knowledge and information commons. While much research has focused on the platform-level administration, few have investigated the actual governance practices of the volunteer leaders whose work sustains the millions of online communities. This dissertation explores those practices through the life cycle of online community rules. Through 40 interviews with volunteer leaders from Reddit communities, Fandom wikis, and Fediverse instances, this research investigates the processes and beliefs behind why and how rules are adopted, enforced, and evolve over time. Findings reveal that rules serve multifaceted purposes, including setting behavioral expectations, reinforcing community identity, and signaling legitimacy to external stakeholders. Enforcement challenges often stem from “grey area” situations, requiring collaborative sensemaking for consistent rule interpretation, and leaders would often strategically build trust to ensure compliance. Rule changes are influenced by leaders’ capacity to address issues and by platform-specific affordances for member feedback. This research advances theories of community governance, content moderation, and institutional change by illustrating the dynamic relationship between leadership decisions, member participation, and the materiality of online community software. Finally, it concludes with implications for designing adaptive sociotechnical systems that support sustainable practices for volunteer-led governance.

TABLE OF CONTENTS

	Page
List of Figures	iii
Chapter 1: Introduction	1
1.1 Dissertation Structure	5
Chapter 2: Theoretical Background	6
2.1 Introduction	6
2.2 New Institutionalism	7
2.3 The Sociomaterialist Perspective	15
2.4 Online Community Research	24
Chapter 3: Empirical Background: Online Communities, Reddit, Fandom, and the Fediverse	31
3.1 Introduction	31
3.2 Reddit	31
3.3 Fandom	43
3.4 Fediverse	60
3.5 Selected Empirical Case Studies and Their Rules	68
Chapter 4: Methodology and Analytical Strategy	77
4.1 Introduction	77
4.2 Data Collection	77
4.3 Analytical Strategy	82
Chapter 5: Rule Adoption: The Functions of Rules in Online Community Gover- nance	85
5.1 Introduction	85
5.2 Setting behavioral expectations	85
5.3 Reinforcing community identity	90
5.4 Signaling legitimacy to external stakeholders	96

5.5	Discussion	101
5.6	Conclusion	105
Chapter 6:	Enacting Community Rules: Compliance, Enforcement, and Trust . .	106
6.1	Challenges for rule compliance: reading the rules	106
6.2	Proactive strategies for rule compliance	108
6.3	Enforcing the rules	110
6.4	Socially constructing rule meaning	123
6.5	Building legitimate authority	128
6.6	Discussion	140
Chapter 7:	Community Evolution: The Processes of Rule Change	148
7.1	Triggers for Rule Change	148
7.2	Supply of Problems	159
7.3	Rule Change Process	161
7.4	Decision-Making Structures	167
7.5	Discussion	171
Chapter 8:	Synthesis and Conclusion	180
8.1	Introduction	180
8.2	Synthesis of Findings and Contributions	180
8.3	Implications for Design and Leadership	184
8.4	Limitations and Future Research	187
8.5	Conclusion	189
Appendix A:	Study Information Sheet	218
Appendix B:	Recruitment Messages	221
Appendix C:	Interview Protocol	225
Appendix D:	Participant Screening Survey	234
Appendix E:	Focused Code Organization	237
Appendix F:	Second Focused Code Organization	239
Appendix G:	Rule Change Process	241

LIST OF FIGURES

Figure Number	Page
3.1	This is a screenshot of the Reddit homepage. Users see a feed of recent posts from the various subreddit communities to which they're "subscribed" in the center of the interface, with the corresponding upvote and comment count. The left side of the interface shows the communities the user is subscribed to as well as other useful information like recent communities visited and moderation information for the subreddits the user is a moderator for. 33
3.2	This screenshot shows an example of a subreddit community's page. Like the Home page, each subreddit has its own feed of posts, showing the number of comments and votes each post currently contains. While the left part of the interface is the same, the right side shows information about the community, including how many members, the number of those online, rules of the community, and any other information the leaders of the community want to include on the right sidebar. 34
3.3	This is a screenshot of the comments on a post in the <code>r/cybersecurity</code> subreddit. While some posts can be media like a video or image, this post is text, with a question asking other cybersecurity professionals about what kind of note taking techniques they use. Under the text post, members of the community reply with comments with their answer. Each comment may also receive an upvote, additional replies (i.e., a "comment tree"), and potentially "Awards," which can be purchased to highlight someone's post or comment and make it stand out in the community. 35

3.4	This is a screenshot of the user interface for Reddit’s “Mod Tools.” Leaders of subreddit communities access this interface to manage and govern the communities in which they moderate. The left side of the interface shows the subreddit communities the user moderates as well as multiple tools that use to facilitate community governance. For example, “Queues” shows the moderator which posts or comments have been reported / flagged for review by other users. Under that is “Modmail”, which is a messaging system for each subreddit to centralize moderator communication to members. Under the “Moderation” category is a page to create a subreddit’s rules, an option to customize what content types are allowed in the subreddit, a log of all moderator actions in their community, and the “Automod”, which is an automated governance tool that works as a bot for automatically sanctioning users and warning them about rule violations.	37
3.5	This is a screenshot of the rules for the r/cybersecurity subreddit. This is not the full list, but this is what users see on the right side of the interface when browsing the r/cybersecurity subreddit feed.	38
3.6	This is a screenshot of a post about a Zero Tolerance Policy rule change in the community made to the r/cybersecurity subreddit by one of it’s head moderators. The post describes the problem and what is going to change . .	39
3.7	This screenshot captures the discussion about the Zero Tolerance Policy rule change by members with the leaders of the r/cybersecurity subreddit. Members can ask questions and give feedback in the comments.	40
3.8	This is a screenshot of the main home page for Fandom. Users can search for fan wikis in the search bar or explore suggested wikis.	44
3.9	This is a screenshot of an example Fandom wiki’s, the Marvel Wiki, home page.	45
3.10	This is an example of an article page on the Marvel Wiki. Members can edit the article, add new content, revert new edits, and discuss aspects about the article on the article’s Talk Page.	46
3.11	Here is a screenshot of a “Talk Page” for the Wolverine article on the Marvel fandom wiki. In this example of a Talk Page, members are discussing possible plagiarism of text in the Wolverine article page.	47
3.12	This is a screenshot of the “Discussions” page in the Marvel fandom wiki. Discussions functions similarly to any other online forum, with users able to add comments and rate posts with “likes.” Not every wiki has a Discussions forum, as it’s an opt-in choice for wiki community leaders.	48
3.13	Here is an example of a post with comments from members in the Marvel fandom wiki.	49

3.14	This is a screenshot of the “Admin Dashboard” for Fandom wiki leaders. Like the Mod Tools on Reddit, this dashboard gives leaders agency over how their wiki is designed as well as how it manages members and member permissions.	51
3.15	This is a screenshot of the default rules that are automatically generated for any newly created wiki on Fandom.	52
3.16	This is a screenshot of the Marvel fandom wiki’s information page on its leaders, titled “Administrators.” On some wikis, like this one, leaders create a hierarchy among themselves. Here, “Bureaucrats” are the head administrators of the Marvel fandom wiki. After them, regular “Administrators” hold power to help manage the wiki and organize the community’s “Moderators”, who are primarily given enough power to moderate content on article pages or in the discussion forums.	53
3.17	This is a screenshot of the list of rule <i>categories</i> (“Policies and Guidelines”) of the Marvel fandom wiki.	54
3.18	A screenshot of the Table of Contents for the first rule category on the Marvel wiki’s Policies and Guidelines Page.	55
3.19	A screenshot of the first policy in the Article Guidelines of the Marvel fandom wiki.	56
3.20	A screenshot of the rules of the Marvel fandom wiki’s Discussions rule set. Note that rules here are more related to interacting with other people and not rules related to formatting article pages.	57
3.21	This is a screenshot of the Talk Page for the Article Guidelines policy section on the Marvel fandom wiki. In this example, members are proposing a rule change about how to format an “info box” on an article, with members discussing whether or not they would be okay with such a change.	58
3.22	This is the interface for a Fediverse instance, sfba.social .	61
3.23	This is an example of how users of Fediverse interact with each other. Note that interactions are similar to those on social network sites like Twitter, where people can comment below a post.	62
3.24	This is a screenshot of the “registration” process on a Fediverse server, sfba.social . Note that potential new members to this community have to read and accept the rules before they can finish signing up. Other registration processes on the Fediverse might be more involved, such as requiring approval from the instance administrator before finalizing account creation.	63
3.25	This is a screenshot of the “About” page of a Fediverse instance, sfba.social .	64
3.26	This is a screenshot of the rules on the Fediverse instance sfba.social . Compared to rules on Reddit and Fandom, rules here seem to be less about formatting and type of content that’s allowed and more about how people behave and interact with other members.	66

3.27	This is a screenshot of the “Moderated servers” on <code>sfba.social</code> . Administrators of Fediverse instances can “defederate” from other communities, which means content can’t be shared across to <code>sfba.social</code> . The reasons for some of the moderated servers here are “Spam problem”, meaning the instance seems to be run by bots or advertisers and “Reports of unaddressed hate speech”, which means members of <code>sfba.social</code> reported the instance for hate speech but it was never addressed by the instance’s admins. Not every Fediverse instance publicly shows which other instances they’ve “moderated.”	67
5.1	Illustrates the three different functions rule may take in the management of online communities.	102

ACKNOWLEDGMENTS

The author wishes to express sincere appreciation to his advisor, Professor Benjamin Mako Hill, for guiding him through years of learning, trial and error, and most importantly igniting his love of social science research and studying online communities. The author also wishes to express appreciation to his doctoral committee members, Professor Aaron Shaw, Kirsten Foot, and Mark Zachry, for their patience and feedback during his doctoral program. The author wants to thank his collaborators who helped with the design, recruitment, interviewing, and analysis of the data in this study: Sohyeon Hwang, Nate TeBlunthuis, Carl Colglazier, and Molly de Blanc. The author also wishes to express his deep gratitude the hardworking and bright undergraduate students at the University of Washington and Northwestern University who helped with the collection and preliminary analysis of the interview data in this dissertation. The author would also like to thank the Community Data Science Collective for accepting him as a member and helping him learn the challenges and rewards of computational social science and open source software. The author would like to express gratitude to Aaron Jiang for his support and cooperation on other projects over the years, as well as gratitude for the few members of the Social Computing Discord server that began in 2018. This work, and much of the author's other research, would not have been possible without the kind volunteer moderators who agreed to give a portion of their time and knowledge in interviews. Lastly, the author would not have been able to pursue this path at all without the unending support and love from his wife, Hillary, who grounded him during the hardest parts of his doctoral journey.

DEDICATION

to the volunteer leaders and earnest participants of online communities whose passion and commitment inspire this work.

Chapter 1

INTRODUCTION

The expansion of human interactions into the networked infrastructure of the Internet produced millions of virtual settlements where communities form, grow, and dissolve. Once frontiers for new forms of communication and collaboration, online communities are becoming increasingly entangled with modern-day social life. They transcend space and time by allowing people to share information and interact freely, regardless of geographic or temporal constraints. The result is the peer production of knowledge and discourse, where people with their own expertise and beliefs can come together to produce a common public good like an encyclopedia (e.g., Wikipedia) or contribute to discussions and collaborative filtering within digital public spheres (e.g., Reddit). However, not everyone that contributes to online communities brings quality or even relevant information, and some can turn civil discourse into toxic arguments. Left unchecked, communities may deteriorate and bleed members as the noise of low-quality content and toxic behavior crowds out good-faith participants. Governance of online communities is therefore an important mechanism for maintaining the integrity and health of these social spaces.

While norms emerge to softly guide participants toward useful contributions, they often have to be encoded as formal rules and enforced to make it costly to disrupt community's goals. The practice of enforcing rules in online communities is frequently called "content moderation," and its efficacy helps to sustain productive and positive experiences on the Internet (Grimmelmann, 2017; Gillespie, 2018). Researchers in computer-supported cooperative work and human-computer interaction studies have sought out effective strategies for managing online communities, but their focus has primarily been on designing the software of the systems and social media platforms that can influence community outcomes (Kraut et al., 2012). Most online communities are not managed by designers of the systems that host them but by volunteer community leaders who have acquired administrative power to

manage membership and regulate the information that’s shared in their community spaces. These volunteer leaders are critical to the success of maintaining civil and productive online spaces (Grimmelmann, 2017; Matias, 2019). An emerging body of work has explored the ways that these groups of volunteer leaders enforce rules (Grimmelmann, 2017; Seering et al., 2019), the challenges they face (Kiene et al., 2016; Wohn, 2019; Schöpke-Gonzalez et al., 2022), and the factors that influence the ways in which community governance structures may evolve over time (Shaw and Hill, 2014; Frey and Sumner, 2019). That said, no work to my knowledge has sought to understand how and why volunteer leaders adopt and change their rules in their communities.

Understanding the practices and challenges that volunteer leaders experience in the actual creation and evolution of their rules has important implications for both designers of sociotechnical systems and for scholars of community governance. A growing concern has emerged that online communities are deterministically pushed toward anti-democratic governance practices because of software design that reinforces power in a small group of individuals in the form of community leaders (Schneider, 2022), and they are concerned that this arrangement will limit the potential for communities to experiment with new forms of governance and evolve (Frey and Schneider, 2021). Empirical evidence about how volunteer leaders decide on how their communities will be governed is critically missing from these arguments. *What factors influence their choices in the rules they adopt, how are the rules interpreted when being enforced, and what factors influence the possibilities for rule change? How are these processes experienced in communities on different sociotechnical systems?*

This dissertation seeks to address these gaps by examining the processes through which volunteer leaders create, enforce, and adapt rules in online communities across three distinct sociotechnical systems: subreddits, Fandom wikis, and Fediverse instances. While prior work has characterized rules on these platforms (Butler et al., 2008; Fiesler et al., 2018; Chandrasekharan et al., 2018; Nicholson et al., 2023), interviews with volunteer leaders provide deep insights into the thought processes, practices, and challenges involved in community governance. By analyzing across these unique sociotechnical systems, I can also begin to determine the extent to which the materiality of online community software infrastructure influences outcomes in community governance.

Through 40 interviews with volunteer leaders across forum communities on Reddit, the peer produced wiki communities on Fandom, and microblogging communities on Fediverse instances, this dissertation reveals complex motivations and practices in the adoption, enforcement, and evolution of community rules. I find that leaders not only adopt rules to regulate behavior and content but also to reinforce community identity and signal legitimacy to external audiences. External pressures from community’s ecological and institutional environments may push leaders to adopt rules that either reinforce their uniqueness or assimilate to broader norms. Rule enforcement also presented challenges, especially in interpreting “grey areas,” and leaders employed strategies to build trust with their members to foster greater rule compliance. Finally, rule changes depended not just on the presence of problems but on also on the capacity for the leaders to initiate the change process, with leaders on Reddit and Fandom more likely to change their rules than those on the Fediverse.

This dissertation makes several theoretical contributions. First, while (Butler et al., 2008) argued that rules help create community identity, my work showed that communities may be pressured into adopting rules that formalize a distinct identity through the adoption of strict rules about participation and contributions. Some communities thrive when they remain a specific size (Hwang and Foote, 2021), and the overlap of communities in topics and members makes for an ecosystem of niche communities where some specialize in member expertise (TeBlunthuis et al., 2022). Rules were a critical component of reinforcing the boundaries between online communities, helping them maintain their identity in periods of existential crisis, like massive influxes of newcomers.

On the other hand, I found evidence of institutional isomorphic pressures on the adoption of similar rules, especially in the decentralized network of communities on the Fediverse. Contrary to an assumption that a decentralized system of authority lacks accountability (Nicholson et al., 2023), this research revealed the coercive, mimetic, and normative isomorphic pressures described by (DiMaggio and Powell, 1983) lead Fediverse communities to adopt and enforce high standards of content moderation and civility. In the absence of a hierarchical, central authority, I argue that horizontal, community-driven forces of conformance may emerge to stabilize relationships in a decentralized community network.

I also extend prior work on the labor of volunteer content moderation (Dosono and

Semaan, 2019; Matias, 2019; Seering et al., 2019; Wohn, 2019) by revealing the extent to which volunteer content moderation constrains the capacity for leaders to govern fairly. While graduated sanctions are a preferred mechanism for fair community governance (Ostrom, 1990; Kraut et al., 2012), I found evidence of leaders abandoning these approaches as they became too costly to enforce due to the scale and nature of the work. I observed leaders using “alt” accounts, a second account for participating as an anonymous non-moderator member, to manage the emotional labor of moderation, protecting their mental health from abuse directed at their moderator accounts.

My work also reveals insights into online community governance as it actually exists in practice. In contrast to arguments that an implicit “feudalism” in the software of online communities creates dictatorial relationships between leaders and their community members (Schneider, 2022), I found evidence of a significant influence of the collective membership on leadership decision-making. Although platforms like Reddit don’t guarantee “effective voice” of members in advancing change (Frey and Schneider, 2021), this research revealed a dynamic relationship where leaders believed facilitating member voice was helpful in providing knowledge about problem solving and establishing trust that went to ensure greater member retention (as well as rule compliance). At the same time, voice and compromise may also have been crucial in preventing regime changes that could arise due to failed perceptions of leadership.

I also make contributions to studies of rule changes in online communities. This work revealed different factors for what may potentially cause a community to change its rules. While prior work found rules were typically added to prohibit types of content (Reddy and Chandrasekharan, 2023), I find that communities may actually remove rules restricting a type of content in order to increase contributions when a community is in decline. I also find evidence of communities never changing their rules at all, as leaders may adopt rules to serve symbolic purposes rather than solve internal problems. I also show how leaders may be aware of problems in their communities but are constrained in their capacity to effectively solve them with a change in the rules, as this adds additional responsibilities to enforce those new rules. Finally, I show how a rule change process can become incredibly complex as leaders navigate how to solve problems, deliberate among themselves and community

members, draft the new rule language, and make a final decision.

1.1 Dissertation Structure

This dissertation is organized into nine chapters. **Chapter 2** describes the theoretical background that influenced the motivation behind the study and situates the work in the broader conversations of institutional change, community governance, and the materiality of software in human organizing. **Chapter 3** provides the empirical background on online communities in general and how governance usually takes place. I also provide detailed descriptions of the three different empirical settings, Reddit, Fandom, and the Fediverse, and summarize the relevant body of work that's explored governance in each. **Chapter 4** describes, in detail, the methodological steps I took in collecting and analyzing data.

In **Chapter 5**, I present the first set of findings on how volunteer leaders adopt rules. The findings here illuminate the functions of rules, from establishing behavioral norms to reinforcing community identity to signaling legitimacy to external stakeholders. **Chapter 6** examines the processes of rule enforcement, challenges and strategies for rule compliance, and how trust and authority are negotiated within these communities. **Chapter 7** describes the last set of findings, which focused on the evolution of community rules. I identify triggers for rule changes and describe the decision-making structures that facilitate the evolution of community governance as well as how these structures varied across the three distinct sociotechnical systems in this study. In **Chapter 8**, I synthesize the findings by connecting them to broader theoretical and practical implications of community governance. I conclude the dissertation by discussing implications for theory and design, limitations, and directions for future research.

Chapter 2

THEORETICAL BACKGROUND

2.1 Introduction

In this chapter, I establish the theoretical foundation and analytical lens used in my analysis on how volunteer leaders of online communities understand, enact, and navigate the governance of their communities with a particular emphasis on the adoption, enforcement, and evolution of online community rules. This chapter serves as a bridge between the empirical findings of the research in this dissertation and the broader theoretical conversations in organizational sociology, new institutional economics, and social computing. To be clear, while not every reference in this chapter directly informs the analysis of interview data in Chapters 5, 6, and 7, this literature review is intended to provide insight into the theoretical perspectives and fields of study that have influenced my own understanding and interpretation of the events and processes described by participants in this study. By situating governance practices of online communities within these bodies of literature, I not only contextualize my findings but also make contributions to research on institutional change in online communities, which are stated in Chapter 8.

This chapter is organized into three sections. First, I describe the “new institutionalism” approach to organizational structure and change. Drawing from organizational sociology and political economy, this perspective helps us understand the dynamics and social forces of rule adoption, enforcement, and change within communities. Second, I describe a complementary “sociomaterialist” approach, which emphasizes the material influence of the software configurations that give online communities a virtual space on their governance outcomes. This perspective highlights the entanglement of technology and social practices, underscoring how the material aspects of technology—such as software infrastructure—shape and are shaped by governance processes in online communities. Through this lens, we can better understand the role of technology design in facilitating or constraining rule adop-

tion, enforcement, and change. Together, these theoretical lenses guided my interpretation of the findings and help us understand the complex relations between social, institutional, and technological factors that mediate governance decisions for online community leaders. In the last section, I highlight the ongoing research and theories on online communities, how they're governed, the practices of content moderation, and the factors that influence online communities' potential for change.

2.2 *New Institutionalism*

In understanding the evolution of governance in online communities, it is essential that we draw upon the rich theoretical foundations provided by organizational sociologists, economists, and political scientists who have contributed to decades of concepts and theories about institutions and institutional change. Theories of institutional development, enforcement, and (in some cases) change offer critical lenses for understanding how rules exist and evolve in the digital realm of online communities. This section outlines the contributions of key scholars—from organizational sociologists to political economists—that inform one side of the theoretical framework, the “new institutionalist” perspective, used in analyzing the findings of the research in this manuscript.

2.2.1 Sociological Foundations of Institutions and Institutional Change

We might begin with Max Weber's concept of rational-legal order as a foundational concept for understanding how modern organizations structure themselves around formalized rules. Weber argued that bureaucracies derived authority from a system of rationally established laws and regulations, and that through this rational-legal order system, society would be trapped in an “iron cage”, with no means of changing course (Weber, 1978). Cyert and March (1963), inspired by Herbert Simon, challenged the assertion that organizations strive toward maximum efficiency, suggesting instead that decision-making of leaders is often messy and driven by a desire to “satisfice”—solutions that are “good enough” but possibly not the most optimal—because leaders are limited in the resources they possess to make perfect decisions (i.e., are limited due by a “bounded rationality”). According to Cyert and March, organizations may change their internal policies and rules when they're triggered

by specific problems in their organization's performance, as if change is always reactive to specific issues that emerge in the organization's history. For a time, organizational scholars expanded on Cyert and March's behavioral theory of the firm, evolving the approach to include theories of ecological dynamics to explain why organizations behave similarly and when there's variation according to factors in their organizational environments (Hannan and Freeman, 1977; Freeman and Hannan, 1983; Carroll and Hannan, 1989).

A new cohort of organizational scholars emerged to challenge the prevailing theories of firms as rational actors. Starting with Meyer and Rowan (1977), they proposed a new line of arguments to explain the behavior of firms. Instead of functioning according to rationalized blueprints for organizations, organizational structures, according to Meyer and Rowan, are often adopted to provide legitimacy rather than directly enhancing their efficiency and performance. Acknowledging that organizations exist within complex relational networks, this approach argues that the formal structures organizations adopt proliferate and diffuse across organizations as highly rationalized myths. In other words, organizations adopt some organizational structure that guides their actions often because by doing so it increases their legitimacy within society, ensuring their survival. Their enactment of formal structure (i.e., the enforcement of rules) may be symbolic so as to maintain the appearance of conformity.

Tolbert and Zucker (1983) built off Meyer and Rowan (1977) by analyzing the diffusion of change in the formal organizational structure of civil service departments. In their work, they found that civil service reforms that begin by sources of high legitimacy, such as the state, diffuse more rapidly and directly to cities than when they begin from no state-level sources, which would diffuse gradually instead. Their findings contribute to the new institutionalist perspective on organizational behavior by providing empirical evidence of organizational change being largely due to the extent to which new policies are institutionalized (Tolbert and Zucker, 1983). In other words, organizations may adopt policies (i.e., rules) when its leaders perceive other, more legitimate organizations adopt those policies first rather than because those policies are adopted rationally to solve actual problems an organization may be experiencing.

A culmination of these ideas emerged in DiMaggio and Powell (1983)'s propositions about institutional isomorphism. DiMaggio and Powell were interested not in why orga-

nizations varied but why they were in so many ways similar. DiMaggio and Powell used the idea of “isomorphism” to describe “a constraining process that forces one unit in a population to resemble other units that face the same set of environmental conditions.” (DiMaggio and Powell, 1983, p. 149) Competitive isomorphic pressures (i.e., market competition, niche change, and fitness measures), had been elaborated on by organizational ecologists (Hannan and Freeman, 1977), but DiMaggio and Powell argued that these only explained part of the process of bureaucratization of organizations. They proposed that another set of mechanisms caused by institutional isomorphism could pressure organizations to conform in similar ways: coercive isomorphism, mimetic isomorphism, and normative isomorphism. Coercive isomorphism emerges from “both formal and informal pressures exerted on organizations by other organizations upon which they are dependent and by cultural expectations in the society within organizations function.” (DiMaggio and Powell, 1983, p. 150) In other words, institutional isomorphic might change might occur when an organization is feeling pressure from the problem of legitimacy or from political influence in their field. Mimetic isomorphism, on the other hand, might result from an organization’s response to uncertainty in their environment, and the pressure results in the simple copying of other organizations’ structures that are perceived to be more successful or legitimate. Finally, normative isomorphic pressures constitute a source of organizational change that is derived from shared professional standards and norms, often driven by formal education, professional networks, and socialization processes within a given organizational field. DiMaggio and Powell’s propositions about these institutional isomorphic pressures as predictors for organizational change are useful for considering why organizations may adopt similar internal policies and rules and what drives them to adopt or change them.

While not strictly aligned with the previously described “new institutionalism” approach, March et al. (2000) explored how the formal written rules of organizations are created, revised, and evolve over time in response to internal dynamics and external pressures. March et al.’s is important for this study because it offers compelling factors to consider as to when and why an organization may change its written rules. For March et al., rules shape and are shaped by organizational events: they are the records of an organization’s history in response to external and internal pressures that jeopardize its goals. March et al. also

provides a working definition for rules: “rules consist of explicit or implicit norms, regulators, and expectations that regulate the behavior of individuals and interactions among them.” (March et al., 2000, p. 5) Rules evolution may occur for several reasons: problem solving, political processes, experiential learning, and diffusion of rules from other organizations.

An important feature of March et al.’s description of the dynamics of rules for this study is the conception of problems, which may be the source of an organization’s change in the rules. Problems may be external such as the emergence of crises, regulations, or norms in their environment or internal, such as increases in the scale of production or activity and increases in the complexity of the organization. That said, March et al. argues that problems are socially constructed—increases in organizational complexity or external shocks might create opportunities but aren’t always causes for problems. For March et al., organizations are “thermostatic in their attention,” meaning that as long as performance is satisfactory, problems may go unnoticed. Patterns of problem recognition are instead produced by the “demand” for any problems at all, rather than their simply being a supply. This conceptualization of problems aligns with Cyert and March (1963)’s concept of “satisficing” but extends it to the phenomenon of rule change. New problems may be perceived as more severe when there’s already a low supply or when a new administrator sees the organization more vividly than their predecessors. Finally, March et al. introduces the idea of the “rule regime”, which is “the system by which rules are created, revised, and maintained, as well as the rule-making agents and processes.” (p. 163) Rule regimes provide us with a framework for understanding how leaders of online communities may be facilitated or constrained to process rule change, as these systems are often coded into the technological infrastructure that supports the online community, such as the way permissions to administer a software package are distributed to some contributors and blocked for others.

This body of work informed my understanding of why leaders adopted rules initially adopted rules in their communities and why they might later change them. For example, in Chapter 5, I cite DiMaggio and Powell’s (1983) concept of “institutional isomorphism” to better understand the social mechanisms that influenced leaders’ decisions to copy rules of other communities. The concept of “satisficing” informed of my understanding of why leaders searched for rules to model that were adequate enough to adopt as their own. In

Chapter 7, I bring March et al.'s (2000) concept of "the supply of problems" and "thermostatic attention" to articulate why leaders did not always change the rules even when problems existed in their communities. I also use the concept of the "rule regime" to help describe situations where community leaders were replaced with new leaders who immediately and completely changed their community's entire rule sets.

2.2.2 The Political Economic Approach to Institutions and Institutional Change

Political economists have generated conceptual frameworks for examining the adoption, enforcement, and evolution of rules in communities. By focusing on the interactions between individuals, communities, organizations, institutions, and their broader social environments, "new institutionalist" political economists provide numerous lenses for understanding how institutions not only constrain and enable individual and group behaviors but also adapt in response to changing economic and political conditions. The scholars discussed in this section are concerned with rules, how rules are successfully enforced, and what factors might lead to their change. This section provides an overview of their concepts and how they contribute to my understanding of institutions and institutional change.

North provided a foundational understanding of institutions, defining them as "the rules of the game in a society, or more formally, are the humanly devised constraints that shape human interaction." (North, 1990, p. 1) He elaborated that these rules "reduce uncertainty by providing structure to everyday life." (p. 1), thereby structuring the exchange of social interactions and goods. Unlike previous political economists, North abandoned the "efficiency view of institutions," arguing instead that decisions within political and economic systems are influenced by incomplete information, cognitive biases, and historical contexts. He suggested that "people decipher the environment by processing information through preexisting mental constructs through which they understand the environment and solve the problems they confront" (North, 1990, p. 20), implying that those in power may not possess an accurate model of the world, affecting the rationality of their decision-making.

Further, North categorized rules as either "informal" or "formal" constraints on actions, referring to social norms and codified laws, respectively. He argued that formal rules are

designed to “promote certain kinds of exchange but not all exchange.” (p. 47), clarifying that rules delineate permissible and prohibited actions, which stabilize social exchanges through reduced uncertainty and enforcement. That said, North also argued that enforcement is imperfect, and he attributes this to the costs involved of enforcing rules and that those in power who enforce rules are often motivated by self-interest, and they may selectively enforce rules or refuse to change them to influence outcomes that benefit them.

North also discussed institutional change, and he says that the sources of change come from shifts in relative prices and tastes. When the benefits of altering the rules outweigh the benefits of maintaining the status quo, those in power may begin a process of institutional change. That said, North argued that these changes are typically incremental because those with the power to change the rules might benefit from the existing, status quo and because the informal constraints of culture and social norms tend to evolve gradually in societies anyway. If positions of power shuffle to new actors, change might be more likely, but even cultural norms survive revolutions (North, 1990). This incremental change is therefore driven by economic incentives and power shifts, which North argued occurs through a path-dependent process where historical trajectories shape the possibilities for institutional evolution.

Political scientist Margaret Levi offers a critical lens for thinking about rule enforcement through the problem of compliance and the relationship between political leaders and citizens. In Levi (1988), rules are understood as being developed as mechanisms for extracting resources and ensuring compliance, driven by the need to balance the interests of the ruling class and their subordinates. Levi explored the conditions under which governments in past societies decide to enforce tax collection and the implications this has for state development and governance. The willingness and ability of a state to collect taxes effectively are critical, Levi argued, for upholding its legitimacy and ultimately its stability. Levi suggested that leaders of these societies make rational calculations about the costs and benefits of enforcement to maximize the effectiveness of collecting taxes from their citizens. That said, even under monarchies, Levi showed how the efficacy of tax collection depended too on the stability of the relationship between the state and the governed—the “social contract”—which needed to be perceived as fair in order to make taxation appear legitimate. Later

work in Levi (2022) argued that compliance to rules is greater when states appear more trustworthy, but this too depends on how well the social contract is upheld. One avenue for building this trust was in the requirement of crafting policies that reflect the values and interests of a state's diverse populations²⁰²². While not as much was theorized about rule change, Levi also argued that changes in policy can provoke opposition from citizens who perceive that they will be adversely affected.

Albert Hirschman's seminal work in *Exit, Voice, and Loyalty: Responses to Decline in Firms* offers a lens for understanding organizational change that can be applied to the problem of institutional change as well. In Hirschman (1970), the concepts of "exit" and "voice" are introduced as strategies that individuals can take to respond to and signal dissatisfaction with the status quo of a product, a political group, or really any organization that provides goods and services for individuals. Exit, according to Hirschman, involves either leaving the organization or abandoning the use of its goods and services, while voice refers to using one's speech to improve or even make aware the deteriorating conditions of the organization, often through communicating directly or a form of protest to mobilize public opinion (Hirschman, 1970). Loyalty influences the choice between exit and voice in which a member or consumer of an organization or product rationalize the decision not to exit. The mechanism of loyalty functions as buffer that can help moderate the inclination to exit a deteriorating organization, giving the organization an opportunity to remedy the situation and enhancing the effectiveness of voice. These concepts are important to consider for institutional change because these same mechanisms can be used to signal to those in power that the current arrangement is performing poorly and that management should do something to change it or risk these individuals withdrawing.

Hirschman also discussed how the presence of competition can significantly affect the mechanisms of exit and voice, potentially making change more or less possible. The existence of competitors might provide an avenue for dissatisfied members to make the threat of exit an effective, viable option for pushing for change, since organizational leaders may want to compromise in order to keep those members from leaving. Conversely, when monopolistic or few alternative and less viable options for exit exist, the effect of exit is less powerful and voice might be the only strategy for signaling that change needs to occur.

This dynamic might play out differently with online communities, in which many, if not hundreds or thousands of alternatives, exist as the initial cost to create an online community is relatively cheap. Competitive pressure to take exit seriously might drive rule change in online communities, but change too might also be limited if effective voice mechanisms don't facilitate this process (Frey and Schneider, 2021).

Finally, Elinor Ostrom's work in *Governing the Commons: The Evolution of Institutions for Collective Action* draws on case studies about the institutional arrangements of communities and the factors that facilitated or limited their ability to change. Ostrom was interested in how communities around the world developed diverse institutional arrangements for managing common-pool resources (CPRs) such as fisheries, forests, and water basins, in order to avoid ecological disasters. Ostrom proposed several "design principles" for successful community-managed commons that include but aren't limited to having rules that match local needs and environmental conditions, participatory mechanisms for changing the rules, graduated sanctions for people who violate the rules, and conflict-resolution mechanisms (Ostrom, 1990). The successful management of a commons relies on community members to collectively ensure compliance via mutual monitoring. Graduated sanctions allows for rule enforcement to be proportional to the severity and context of the rule violation, which helps foster cooperation in the community because this approach would be perceived as fair and equitable.

Like North (1990), Ostrom also argued that institutional change is incremental rather than radical and perceived it through a cost and benefit analysis. Ostrom states out that radical change is costly—one must supply an entirely new institution. As Ostrom points out, "if individuals find rules that work relatively well, they may have little motivation to continue the costly process of searching for rules that will work even better." [p. 211] (Ostrom, 1990) Instead, refining rules over time is the more common path communities take as it introduces less uncertainty than radically changing them at once. Ostrom argued that individuals would be more willing to adopt new rules, especially if they restricted their actions of the CPR, when the CPR is deteriorating or when leaders of a community are successful in convincing others of an impending crisis (Ostrom, 1990). Ostrom's design principles and conceptualization of self-governing communities' institutional arrangements

are influential in studies of online communities, as many are formed around their own CPR (e.g., a peer-produced encyclopedia like a wiki or open source software project). Ostrom’s insights reveal that the sustainable governance of a commons is largely contingent on adaptable, locally-tailored institutions that evolve through community engagement and pragmatic incrementalism.

The political scientists and new institutional economists discussed in this section were important in informing my understanding of rules as institutions, the challenges of rule enforcement, and rule change. For one, I use North’s (1990) definition of rules to ground my understanding of what rules do in online communities in every chapter of this dissertation. Rules provide the formal framework for the norms and behavioral expectations that guide the ways that people in online communities interact with each other. They are not only symbolic of online community governance—their enforcement is what helps community strive toward their goals and achieve stability. I cite Levi’s (1988) work on “quasi-voluntary compliance” in Chapter 6 to help articulate the trade offs that online community leaders, who are in positions of authority, experience to maintain legitimacy. Levi’s (1988) work gives us a foundational understanding of why leaders compromise and why trust is important in maintaining stability community governance. In Chapters 6 and 7, I use Hirschman’s (1970) concepts of “exit” and “voice” to articulate the relationship between leaders and members of online communities and the ways in which members can pressure leaders to solve problems in their communities by changing the rules. I also reference Ostrom’s (1990) work on the governing of common pool resources in Chapters 6 and 7. The management of online communities is similar to that of common pool resources found in Ostrom’s (1990) research because, left unmanaged, these spaces can deteriorate and be less useful for everyone overall. I take Ostrom’s understanding of enforcing and changing rules as being costly to inform my understanding of the individual decisions that leaders take when solving problems in their communities.

2.3 The Sociomaterialist Perspective

The previous literature considered the evolution of governance for organizations and communities that physically exist in the world, but online communities, the empirical cases in

this study, exist virtually on the Internet. While governance in the physical world, despite having material effects, is structured abstractly through documented systems of organizing, governance for online communities is entirely mediated by the the technology of the Internet and the software that gives digital “space” to Internet users. Software code can effectively regulate behavior of actors in online communities as it limits what actions are permissible and which are prohibited (Lessig, 2006), and this code can be embedded with implicit ideologies favoring a single kind of governance structure (Schneider, 2022). Software configurations allowing all users to make changes could promote democratic forms of community governance, while a different software configuration centralizing control to a single administrator could lead to autocratic community governance. Along with the social arrangements that influence the decisions of governance in online communities, we must consider the influence of technology design on the outcomes observed in this study. In this section, I review the literature theorizing the relationship between technology and organizing. Drawing on scholarship from social construction of technology, computer-supported cooperative work, organizational communication, and computer-mediated communication, I present the key theoretical lenses that make up the “sociomaterialist” perspective that informed my analysis of the data.

2.3.1 Science, Technology, and Society Studies

The interdisciplinary field of science, technology, and society (STS) studies sets the foundation for forming an understanding between the relationship between technology, society, and the role of human agency. One of the most important concepts to come out of this field is the “social construction of technology,” which builds off Berger and Luckmann (1966)’s concept of “social construction”, the idea that reality is not inherently objective or fixed but instead created through social processes and interactions among individuals within a society. In a similar sense, the theoretical framework of social construction of technology (SCOT) argues that artifacts (i.e, technologies) are ascribed various meanings by different groups due to their “interpretive flexibility.” (Kline and Pinch, 1996; Bijker, 2010) Scholars of SCOT challenge the assumptions of technological determinism and show how users of

technologies influence both their design and use and ascribing to it new meanings. For example, Kline and Pinch (1996) showed how the introduction of the automobile was more than a transportation device but as a source of power that reinforced gender roles in rural America. However, just as social constructions become institutionalized through consistent reinforcement, technologies too may gain a kind of momentum that embeds these meanings into society, making them harder to change (Latour, 1990; Bijker, 2010).

Another problem for STS scholars is the extent of their neutrality and agency. There is a tendency to blame negative outcomes on technologies with the technological design itself, but the meaning and uses of technologies cannot be separated from the people and activities it influences (Pacey, 1983). A critique of SCOT emerged when Winner (1980) punted an argument that technologies are also inherently political. Technologies might be developed to reinforce certain forms of social order, such as the design of McCormick's reaper in breaking the power of labor unions in the United States (Winner, 1980). Some technologies, he argued, "are by their very nature political in a specific way...the adoption of a given technical system unavoidably brings with it conditions for human relationships that have a distinctive political cast." (Winner, 1980, p. 128) Winner's conceptualization of embedded politics in technology is not necessarily deterministic, but it acknowledges their materiality in a way that might be overshadowed by SCOT scholars. For a relevant example to this study, Schneider (2022) showed how social computing systems have historically been designed in such a way that reinforces anti-democratic political arrangements for online communities. Although not always explicitly cited in the finding sections, STS theories permeated my understanding of how the technologies that host online communities may be imbued with specific ideologies or hierarchies that influence the ways that leaders in the communities I studied make decisions about how to govern.

2.3.2 Technology and Organizing

Building off the above foundational understanding of the relationship between technology and society, this section reviews the literature on how technology shapes and is shaped by organizational structures and work practices. Scholars from organization science, computer-

supported cooperative work, and human-computer interaction responded to the growing work on the social construction of technology by pushing back with an emphasis on the materiality of technology in influencing human actions. This body of work helps to define and underscore the “sociomaterialist” perspective that helps us consider how the different software configurations studied in this research facilitate, constrain, and reinforce different governance outcomes and institutional change.

A core theory influencing the scholarship in this section is Anthony Giddens’ “structural theory”, which essentially argues that society embodies a duality of both social practices and social structures (Giddens, 1984). Earlier work by Barley (1986) gathered empirical evidence showing that technologies do not deterministically shape organizational structures but rather act as a trigger or “occasion” for structuring (and potentially organizational change). Barley argued that structure is an ongoing process between technology, human actions, and institutional frameworks. This work showed how the introduction of a new technology used in similar contexts could result in different organizational structures because of the influence of existing social orders and frames for interpreting how technology should be used.

Orlikowski (1992) expanded on this work by rejecting the notion of technology as either a deterministic force or a mere outcome of human agency. Instead, she introduced the concept of the “duality of technology,” highlighting its capacity to both shape and be shaped by human actions. In her study on the adoption of *Lotus Notes* in the 1990s, Orlikowski introduced the idea of “technological frames”—the assumptions, knowledge, and expectations that users bring to their understanding of technology (Orlikowski and Gash, 1994). Drawing on concepts of interpretive flexibility from STS and cognitive framing from Erving Goffman, Orlikowski argued that these technological frames influence how users make sense of technology, often leading them to reproduce existing organizational structures and processes in their use of new technologies.

While structuration theory influenced studies of technology in organizational practices, Activity Theory was introduced into human-computer interaction studies to critique the traditional cognitive approaches that are often too narrowly focused on internal human processes instead of broader social and cultural contexts (Kaptelinin and Nardi, 2007). This

framework emphasized the importance of understanding human actions as being mediated by technologies and artifacts in the context of human activities in what is called an “activity system” (Foot, 2014). Technologies (tools) help individuals (subjects) achieve their goals (objects), and each node in the activity system influences the other. Activity Theory has parallels to Latour’s Actor-Network Theory in that it treats humans and technology not as separate entities but as interconnected elements within a network of relationships, where both human and non-human actors play active roles in shaping outcomes and driving processes within an activity system. This perspective challenges the notion of technology as merely a passive tool, instead highlighting its agency in co-constructing social practices and influencing the dynamics of human interaction. Organizational approaches to studying technology and organizing were important in how I understood the ways in which online community leaders think about governing *through* technologies. This body of work helps me situate online community governance in the actual, everyday practices that leaders discussed in interviews.

Sociomateriality

Building off decades of research on technological change in organizations, Orlikowski and Scott (2008) presented the concept of “sociomateriality” to help researchers of technology and organizing to move beyond traditional separations between the social and the material. Sociomateriality argues that the social and material are inextricably intertwined in organizational practices and that technology, work, and organization not only should not but cannot be studied as separate entities. Their concept emerged in response to decades of organization studies that treat technology as if it’s “absent from the world of organizing, despite its pervasive presence in organizational work.” (Orlikowski and Scott, 2008, p. 433) Technologies are largely ignored or taken for granted or studies focused mostly on the adoption of technology rather than treating it as a constant aspect of organizing work practices. According to Orlikowski, “the social and the material are considered to be inextricably related—there is no social that is not also material, and no material that is not also social” (Orlikowski, 2016, p. 1437). In other words, organizational practices are sociomaterial prac-

tices, and ignoring materiality leads to a limited understanding of organizational life and its consequences.

The sociomaterialist framework is useful for understanding organizational change. Orlikowski (1996) uses sociomateriality to understand change in organizations as an ongoing, improvisational process embedded in everyday organizational practices. The role of technology from this perspective situates it an autonomous, ongoing force that helps drive organizational change rather than view change as a deliberate, radical action by managers. Leonardi and Barley (2008) also emphasized the role of materiality in shaping how technology is used within organizations and how organizational practices evolve as a result. They argued that future research should develop typologies of constraints and affordances that technologies provide to better predict how technology might impact work and organizational structures. Sociomaterialism helps develop more robust understandings of organizational change because it situates change in every day practices of organizational actors who adapt and innovate in response to contingencies and breakdowns in their work (Orlikowski, 1996). For my work, the focus on everyday practices of leaders of online communities in adopting, enforcing, and changing rules aligns with the sociomaterialist perspective by considering how either communities built on similar software infrastructure arrive at different governance structures or communities built on different software infrastructure arrive at similar or different outcomes.

The sociomaterialist perspective guided my understanding for how these community governance dynamics play out in different digital environments, especially in my discussions in Chapters 5-7. Considering the design of the systems that host the online communities in my study might lead us to believe that certain governance outcomes are highly likely. For example, peer production communities on Fandom wikis are built on software that encourages open participation and contributions. This might mean that governance, too, involves encouraging participation of all members to help decide on rules or when to change them. Subreddits, on the other hand, exist in software infrastructure that structures a hierarchical leadership around the creator of the subreddit, its moderators, and then non-members, who have no administrative permissions. This arrangement may encourage autocratic or oligarchical governance outcomes because of Reddit's materiality.

2.3.3 *Organizational Communication*

My analysis is also influenced by scholars from the field of organizational communication. Organizational communication is a field that intersects human communication and organization studies. Organization studies had historically prioritized the “managerial voice” with normative claims about communication used as strategies to achieve efficiency outcomes for firms (Mumby and Stohl, 1996). Instead, organizational communication scholars should see communication as “the substance of organizing in the sense that through discursive practices organization members engage in the construction of a complex and diverse system of meanings.” (Mumby and Stohl, 1996, p. 63) In other words, communication plays a central role in the process of organizing and through communication, the sense of what the organization is and means is built and rebuilt through ongoing communication with members. Taken to a more extreme approach, organizational communication scholars may draw on the theory of communicative constitution of organizations. This approach understands organizations as an emergent network of self-referential decisional communications where past decisions become the “premise” for future decisions, perpetuating the organization’s existence (Schoeneborn and Vásquez, 2017a). The organization is an assemblage of its identity, narrative, and structure that’s constituted and reinforced through organizing (actions, processes, and technologies). Together, these approaches problematize the concept of the “organization”, decoupling the myth of its identity from the actions that perpetuate it.

An important concept emerging out of organizational communication is “sensemaking.” Sensemaking “involves the ongoing retrospective development of plausible images that rationalize what people are doing” (Weick et al., 2005, p. 409). In other words, Weick et al. argued that people in organizations look back at past events to make sense of the present situation and construct a narrative that can guide what actions to take in future decisions. These processes are inherently social and involve collaborating with other organizational members to build a shared understanding, with communication central to this process. In the context of community governance, organizational communication scholars would consider sensemaking as a critical part of how communities are organized, how rules are developed, enforced, and when problems become severe enough to warrant a change. This

concept also reinforces the idea of problems (i.e., problems that could be solved with a change or addition to the rules) as being socially constructed (March et al., 2000). Sense-making is a useful theoretical concept for understanding how power dynamics influence the interpretations of situations in the sensemaking process, potentially shaping governance outcomes that benefit them or contributing to the institutionalization of certain interpretations (Weick et al., 2005). Theories of organizational communication were important in guiding my understanding of the processes that leaders took in describing how they made sense of their rules when enforcing them in Chapter 6 and the ways in which they made sense of discussed problems that led to rule change in Chapter 7.

2.3.4 Computer-mediated Communication

While organizational communication scholars aim to understand organizational outcomes through the analysis of the everyday communicative practices of firms in the physical world, computer-mediated communication scholarship brings back the problem of materiality by addressing how communication is impacted by computer-mediated technologies. Computer-mediated communication is therefore an important body of work to consider when attempting to understand how groups who come together and interact entirely through the digital, networked infrastructure of the Internet can organize effectively at all and how the entirely technologically mediated setting affects their work.

To summarize the field, computer-mediated communication (CMC) differs from face-to-face communication because it diminishes nonverbal social context cues expressed through the face, body, motion, voice, etc. (Walther, 1996). In CMC settings, communication may become hyperpersonal as individuals imagine idealized perceptions of the person sending them information and simultaneously present curated versions of their own self-image (Walther, 1996). CMC might entail asynchronous communication that allows individuals to interact with each other across time, allowing tasks and social exchanges to occur without time pressures and can improve the efficacy of communication, allowing for “planned discourse” (Walther, 1996).

Affordances

Modern approaches to CMC scholarship stress the importance of studying the mediation of communication and less on the technological object itself (Carr, 2020), and one concept useful for understanding the mediation of communication by CMC technologies is affordances. The concept of affordances emerged out of psychology in the 1960s and referred to the properties of the environment that offer potential actions to an organism (Gaver, 1992). In other words, for CMC technologies, affordances means “the multifaceted relational structure” between an object or technology and a user that enables or constrains potential behavioral outcomes in a particular context (Evans et al., 2017). Affordances are socially constructed—they exist as relational constructs sitting between objects (technologies) and outcomes, influencing but not determining those outcomes (Evans et al., 2017). For example, in social media, editability is an affordance because it allows for users to craft and later revise messages before and after they’re shared, giving them control over how their communication is presented and perceived by other users (Treem and Leonardi, 2013).

Treem and Leonardi (2013) describe four affordances of social media that enable unique communication behaviors in organizations that weren’t possible with traditional CMC technologies: editability, visibility, persistence, and association. Visibility is an affordance that’s enabled by social media by the sheer act of allowing information to be shared and communicated across organizational boundaries. Persistence can be achieved when social media enables communication to remain accessible after initial publication, such as how wikis persistently show articles and subreddits persistently show discussions between users. Association enables the formation of connections between individual users, individual users and content or between pieces of content shared on social media, which can facilitate the creation of social networks and knowledge sharing. Affordances can significantly alter organizational processes like socialization, knowledge sharing and power dynamics (Treem and Leonardi, 2013). The concept of affordances might not just be applicable to technology but also the possibilities for governance. For instance, we might consider how different social media sites hosting online communities enable or constrain participation in decision-making processes like changing rules or participation in enforcing them on other members. In the

next section, I will present theories and research about online communities and how they enact and experience governance and change.

Theories from computer-mediated communication helped deepen my analysis of the ways in which the differences between the sociotechnical systems in my study influenced the opportunities and constraints for community governance change. Specifically, in Chapter 7, I take the concept of “affordances” to explain how and why leaders from Reddit seemed to incur more community-led instances of rule change compared to rule change instances on Fandom and Fediverse communities. Together with theories on sociomateriality, CMC scholarship grounded me in my considerations for how the design of the software technologies that hosted these communities may facilitate or create friction in allowing community governance to evolve.

2.4 Online Community Research

This section introduces the concept of the “online community” and reviews the body of social computing research that investigates the mechanisms behind community governance and change. By defining “online community” and outlining its typical governance structures, I provide a foundation for understanding the empirical contexts of my study. Communities may be generally defined as “identified groups of interdependent people who discuss actions and share certain practices and have a concern for the common good.” (Bellah et al., 1985, p. #). Communities exist everywhere in virtually every society, from academic fields to identity groups (e.g., the LGBTQ community) to fandoms of popular media and so on. The notion of a community existing on the Internet has been used a way to study how groups who may be decentralized and spread out geographically, and often interact asynchronously, come together and thrive through computer-mediated communication. The concept of an “online community” emerged in the time of the early Internet when users would access bulletin board systems (BBS) and email list servs and Internet relay chats (IRC) (Jones, 1997; Schneider, 2022).

One of the earliest definitions came in the form of the “virtual community,” which Rheingold (1993) defined as “social aggregations that emerge from the Net when enough people carry on those public discussions long enough, with sufficient human feeling to form

webs of personal relationships in cyberspace.” Jones (1997) introduced the idea of the “virtual settlement” to describe the conditions under which an online community may form: “1) a minimum level of interactivity, 2) a variety of communicators, 3) a minimum level of sustained membership, and 4) a virtual common-public-space where a significant portion of interactive group-CMC’s [computer mediated communication] occur.” The concept of the “virtual community” matured to include, along with the above conditions, “a shared purpose administered by guidelines and policies within a computer system” (Preece, 2000) With some semblance of governance and rules entered the concept of the online community. Leimeister et al. (2004) offered a more robust definition:

A virtual community [also known as an online community] consists of people who interact together socially on a technical platform. The community is built on a common interest, a common problem or a common task of its members that is pursued on the basis of implicit and explicit codes of behavior. The technical platform enables and supports the community’s interaction and helps to build trust and a common feeling among the members.

This definition allows me to operationalize what I mean by “online community” in this study. Online communities don’t have to be places of discussion—they may also be in pursuit of a “common task,” like the peer production wikis on Fandom. That said, it’s more useful, as Bruckman (2006) argues, to pursue different categories of communities instead of trying to fit them all under a prototypical community definition. My study examines three kinds of online communities: centralized discussion communities (subreddits), peer production communities (Fandom wikis), and federated microblogging communities (Fediverse instances).

2.4.1 Community Success

While online communities may emerge to accomplish various goals or for different reasons on different platforms, their long-term success often hinges on effective management and decision-making to navigate the challenges they face and facilitate a thriving community environment. In many cases, and in those in my study, online communities begin when one

user on a platform creates a community space (e.g., a subreddit, a Fandom wiki, or a Fediverse instance). One of the biggest challenges of a community taking off and succeeding is in attracting and retaining new members (Kraut et al., 2012). Additionally, communities that don’t resolve problems may risk failing to retain members and attract new ones, potentially resulting in community “death” (Iriberri and Leroy, 2009).

Social computing scholars have been interested in the factors that contribute to community success for several decades. For examples, Butler (2001) studied the role of size and communication activity in sustaining online communities over list serves, finding that either factor can contribute to its success, but ultimately needing a balance of growth of activity. Work by Farzan et al. (2011) found evidence in support of the possibility of designs in the systems that host online communities in facilitating member commitment and retention. Other work analyzed the effect of turnover—the churning of old members leaving and new members joining—on online community success, finding that it’s beneficial because new members increase the perception of liveness in a community and increase levels of participation all around (Dabbish et al., 2012). Preece et al. (2004) showed how community members may not contribute publicly (i.e., “lurkers”) but are still crucial to their success. Leadership of online communities has also been studied by social computing scholars. Zhu et al. (2013) analyzed the effectiveness of shared leadership in sites like Wikipedia to understand how collective outcomes are achieved. A large statistical analysis by Kraut and Fiore (2014) evaluated the success of online communities based on attributes about their founders, finding that more experience with online groups and more resources to newcomers (e.g., online friends) contributed to higher levels of success.

2.4.2 *Governance and Moderation in Online Communities*

Scholarship on factors contributing to the success of online communities have resulted in a series of design recommendations (Kraut et al., 2012). In recent years, a range of scholars have sought to understand how online communities govern themselves *despite* the designs of the systems that facilitate their communities (Frey et al., 2019; Kiene et al., 2019b; Cai and Wohn, 2021; Anaobi et al., 2023; Jhaver et al., 2023a; Kuo et al., 2023). This shift in

focus has led to a deeper exploration of the governance mechanisms that emerge organically within online communities, beyond the influence of platform design. As a result, I turn to the research on how governance is enacted and experienced in online communities.

The term “governance” describes the social practices, activities, and processes that shape behavior within and between institutions (Bevir, 2012). Governance involves both formal rules and laws as well as informal norms and actions taken by those with power. Prior work on governance in online communities has, for example, explored the ways in which policy-making and democratic deliberation may take place (Black et al., 2008), the functions of rules (Butler et al., 2008), how members handle conflicts and resolve disputes (Billings and Watts, 2010), and factors that influence political arrangements like oligarchies (Shaw and Hill, 2014).

Decades of research in social computing has sought to reveal the ways in which online communities are governed. Self-governance is a particularly important feature of online community platforms where users can create new online communities and then recruit other members to help manage them. These online community managers, often called volunteer content moderators, are tasked with screening content, such as images, links, and text (Gillespie (2018). Although moderation work has been characterized as custodial in nature (Gillespie, 2018; Seering et al., 2022b)—and most accounts agree that much of this work involves uncontroversial acts of cleaning up—moderation in online communities necessarily involves deciding what is or is not allowed and is an essential part of what (Grimmelmann, 2017) describes as “the governance mechanisms that structure participation in a community to facilitate cooperation and prevent abuse” (Grimmelmann, 2017). A growing line of work on moderation in social computing scholarship has documented the challenges and experiences of volunteer moderators (Dosono and Semaan, 2019; Wohn, 2019), and has sought to understand how volunteer moderation teams are internally organized (Seering et al., 2019), how they engage in collective action (Matias, 2016b), and how they deal with change (Kiene et al., 2019b).

Because governance is, at least to some degree, determined by these moderators, the diversity of volunteer moderators across a population of self-governing communities allows for the possibility of diversity in governance. What the moderators of one community

forbid might be tolerated or even celebrated in another similar community within the same platform. For example, separate studies by Fiesler et al. (2018), Chandrasekharan et al. (2018), and Nicholson et al. (2023) have each documented an enormous range of policies created by volunteer moderator teams in platforms like Reddit and Mastodon (a Fediverse platform).

2.4.3 Community Evolution and Change

While early social computing research oriented toward factors driving community success, an alternative perspective that online communities have “life cycles” suggests that not every problem can be designed away. Iriberri and Leroy (2009) proposed a life cycle framework of online communities in which they generally follow a similar trajectory: inception, creation, growth, and maturity (or death). Each stage of the life cycle may require not only different technologies but also different management strategies to survive in their environments. Online communities often strive to grow to some extent, as attracting, retaining, and turning over members helps communities stay active (Farzan et al., 2011; Dabbish et al., 2012; Kraut et al., 2012). However problems may emerge from growth (Kiene et al., 2016), and changes may be necessary to community governance that ensure the community’s survival (Iriberri and Leroy, 2009; Lin et al., 2017). Online communities may fail (Ducheneaut et al., 2007), and ignoring organizational problems and refusing to evolve may contribute to their demise (Kiene et al., 2018).

A small body of work has explored the evolution of governance in online communities. Earlier work by Butler et al. (2008) analyzed the ways in which rules evolve on Wikipedia. Seemingly open and egalitarian peer production sites like Wikipedia were found to develop bureaucracies for controlling the quality of contributions. Rules generated more rules, including rules about how to add or change the rules. Later work by Shaw and Hill (2014) took sociologist Robert Michels’ concept of the “iron law of oligarchy” and tested it empirically on thousands of wiki communities. They According to Michel’s theory, voluntary organizations that grow in size and complexity may turn oligarchical as a small number of early members “consolidate and exercise a monopoly of power within the organization as

their interests diverge from the collective's." (Shaw and Hill, 2014, p. 216) In other words, these scholars were interested in how online communities that start off relatively democratic and open become less democratic over time. Peer production sites often *feel* democratic because their success often hinges on allowing anyone to contribute to and take care of the public good, information commons. However, Shaw and Hill found that even wikis may change when their community membership grows, as a small group of early contributors "monopolize positions of formal authority in the community and accounts for more administrative activity while also using their authority to restrict contributions from experienced community members" (Shaw and Hill, 2014, p. 229) In other words, an online community situated in a technical setting that could and should (in theory) facilitate and encourage democratic forms of governance was instead shaped by a range of social pressures, leading to less democratic outcomes.

The type of governance structure of an online community may influence its potential for change. Schneider (2022) showed how the technological infrastructure supporting online communities may be imbued with an implicit form of organizing resembling feudalism, and that this implicit ideology guides and limits users from experimenting with more participatory forms of governance. Frey and Schneider (2021) suggested that communities may require mechanisms for "effective voice" to drive change, and in another study, argued for a "constitutional layer of stakeholder participation" so that conflicts in online communities could be more successfully resolved (Frey et al., 2019). This change could be in the form of the rules, or it might also be changing those in positions of authority. Other work has sought out factors that drive community evolution. For example, Frey et al. (2022) analyzed communities on *Minecraft*, the MMO game *World of Warcraft*, and Reddit and found that the complexity of their governance was positively associated with the community's maturity in terms of community activity, age, and size. Work by Reddy and Chandrasekharan (2023) analyzed the evolution of rules of subreddits and found that a majority of changes happen only within the first year of a community's creation, and additional changes to rules will typically correspond to behavioral problems that emerged in the communities and had to be resolved. While a community's maturity tends to be a significant factor in community change, my study expands on this work by examining the individual human actions of lead-

ers as well as the affordances and constraints of community governance technology in the capacity for community change.

This chapter has outlined the theoretical background that has influenced my understanding of how and why the volunteer leaders of online communities who participated in this study navigated governance in their communities. Drawing on new institutionalism, sociomaterialism, and scholarship on online community governance, I have described how social, institutional, and technological factors can potentially shape the processes of leaders of online communities. With this literature review, I have helped situate the reader within the broader discussions of rule adoption, enforcement, and change in online communities and established the foundations on which this work has made several contributions. In the next chapter, I will describe in more detail the specific platforms that hosted the online communities in this study, the types of communities from which we sampled volunteer leaders, and the kinds of rules that these communities typically adopt.

Chapter 3

EMPIRICAL BACKGROUND: ONLINE COMMUNITIES, REDDIT, FANDOM, AND THE FEDIVERSE

3.1 *Introduction*

This chapter introduces the specific empirical settings, Reddit, Fandom, and the Fediverse, where we selected cases for online community governance and interviewed volunteer community leaders. I provide detailed descriptions of the user interfaces and experiences within these settings, grounding the reader in the technological environments under examination. I review previous research on these platforms to situate our findings within the broader scholarship on governance of online communities, setting the stage for the contributions discussed in later chapters. Finally, I summarize the kinds of rules present in the communities from each sociotechnical setting and describe the similarities and differences observed.

3.2 *Reddit*

Reddit is a website that also allows users to create their own spaces for communities called “subreddits”. Subreddits are functionally similar to forums where one person creates a post on which others can then comment and discuss. All posts and comments, however, can be given a single “upvote” or a “downvote” by each user. This voting system is used to facilitate community curating of content—posts and comments with more upvotes are generally the first to be seen on a subreddit’s home page or within the post’s comment section. Membership to communities on Reddit is somewhat ambiguous. Users create an account on Reddit and can “subscribe” to a subreddit community, and I use the total subscriber count as a measure of community size. That said, actual participation is not the same as the subscriber account, and users can subscribe to as many communities as they desire. Leaders of subreddit communities are volunteers and are generally recruited from the community’s membership base. While Reddit enforces sitewide policies through

site administrators, subreddit communities are tasked with enforcing their own governance structure over their community spaces.

Figure 3.1 illustrates the user interface for a user of Reddit. Reddit makes it relatively simple for users to find communities, either through the search bar or by finding communities that pop up on their “feed”, the stream of posts from communities across Reddit. Once on an individual subreddit, like `r/cybersecurity` in the screenshot in Figure 3.2, a user can “join” the community (also called “subscribe”), which will make it so new posts from that community show up on the user’s “Home” feed. Each subreddit community consists of posts, usually by other users who have “joined” the community, and those posts can be commented on and upvoted, as is illustrated in the screenshot in Figure 3.3. The voting system helps filter relevant and / or quality information in the subreddit, which can be easily accessed by sorting posts (or comments) by their number of upvotes.

3.2.1 Reddit Rules and Governance

Like most online communities, subreddit communities are created and staffed entirely by volunteer leaders called “moderators.” A subreddit is created by a Reddit user (i.e., “Redditor”) and automatically becomes a moderator for their subreddit. That user may add other Reddit users as moderators to assist them in managing their community space. Reddit supplies community moderators with a collection of “Mod Tools”, which is shown in the screenshot in Figure 3.4. Subreddit moderators can track sanctions that have been issued against rule violators in their communities as well as reports about suspected rule breaking posts or comments. Tools also allow moderators to manage the users in their communities, communicate to members semi-anonymously through the “Mod Mail” system, and create rules and enforce them automatically with the “Rules” and “Automod” feature. Rules may be made visible on the right side of the user interface for anyone browsing their subreddit, like I show in Figure 3.5. While the Rules feature only allows 15 rules maximum to be listed on a subreddit, moderators have utilized the subreddit “Wiki” feature to add an additional page with more information and FAQ about the rules. Subreddit moderators frequently interact with their community members, especially through posts that may or may not be

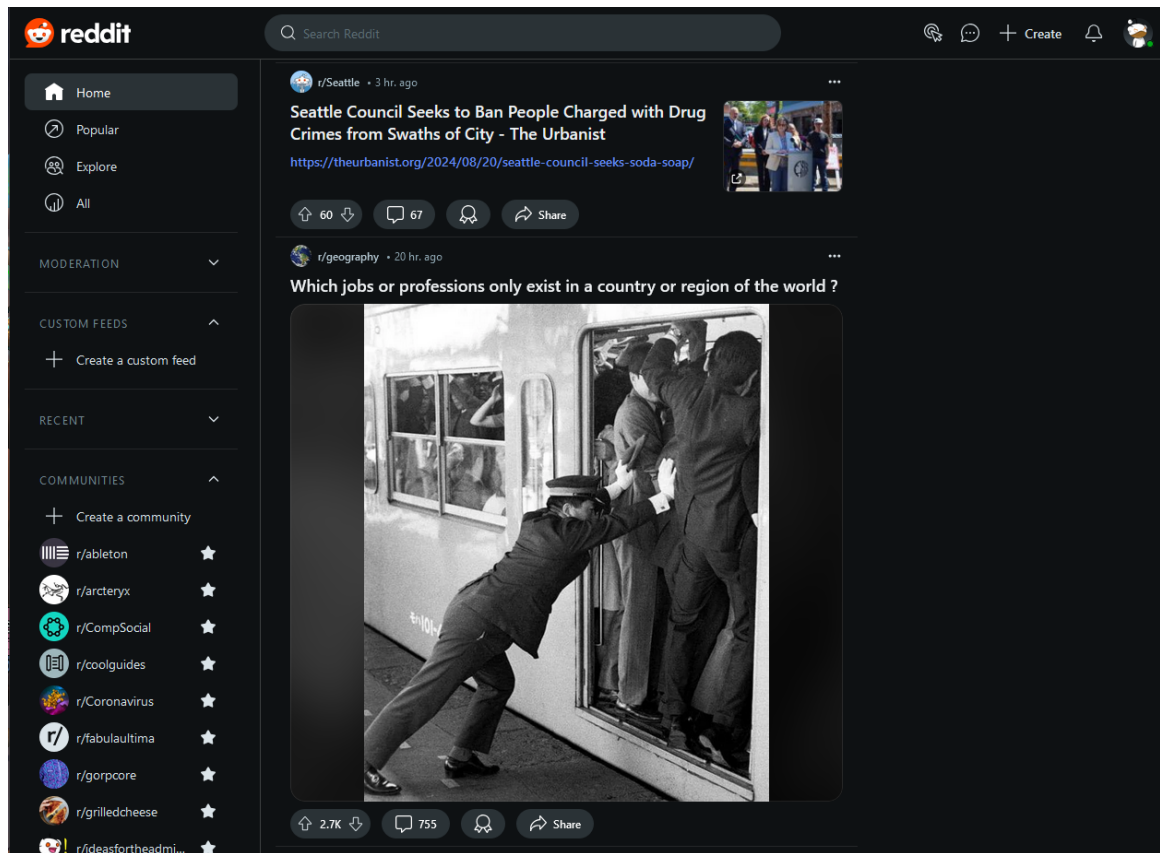


Figure 3.1: This is a screenshot of the Reddit homepage. Users see a feed of recent posts from the various subreddit communities to which they’re “subscribed” in the center of the interface, with the corresponding upvote and comment count. The left side of the interface shows the communities the user is subscribed to as well as other useful information like recent communities visited and moderation information for the subreddits the user is a moderator for.

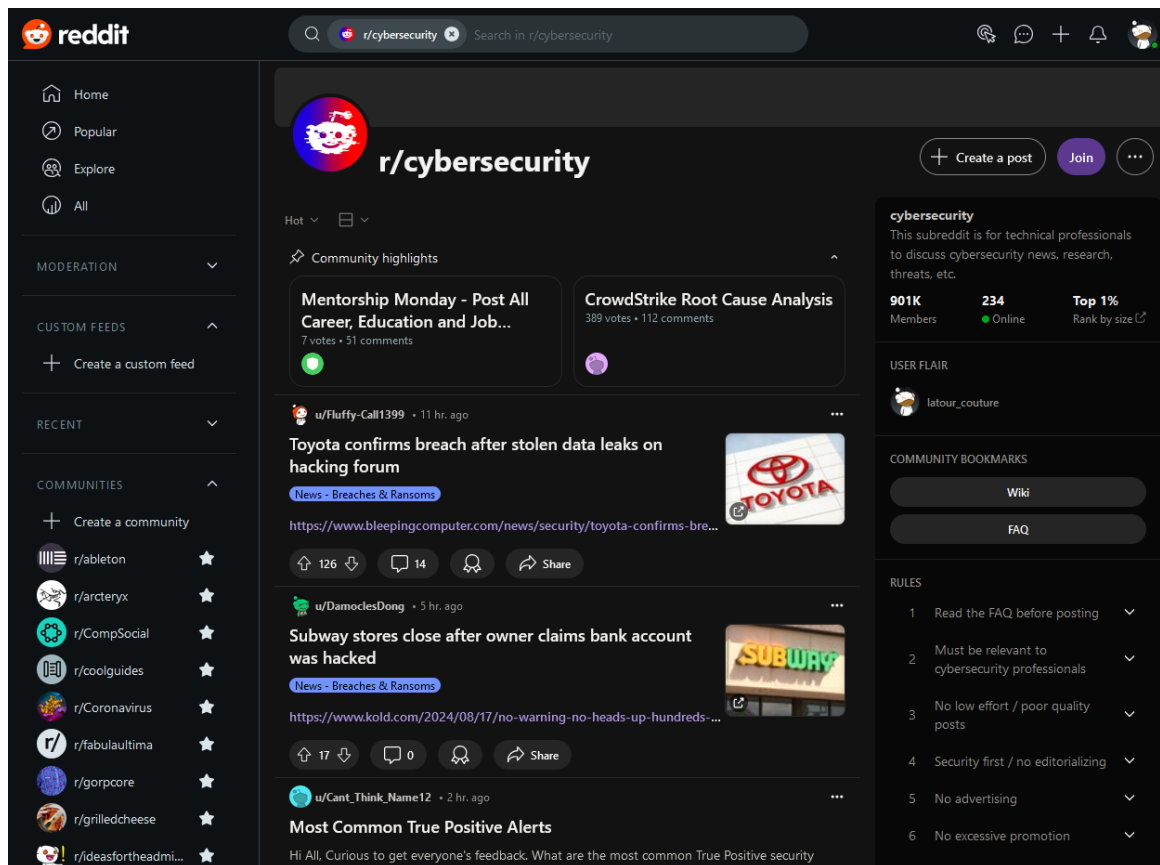


Figure 3.2: This screenshot shows an example of a subreddit community's page. Like the Home page, each subreddit has its own feed of posts, showing the number of comments and votes each post currently contains. While the left part of the interface is the same, the right side shows information about the community, including how many members, the number of those online, rules of the community, and any other information the leaders of the community want to include on the right sidebar.

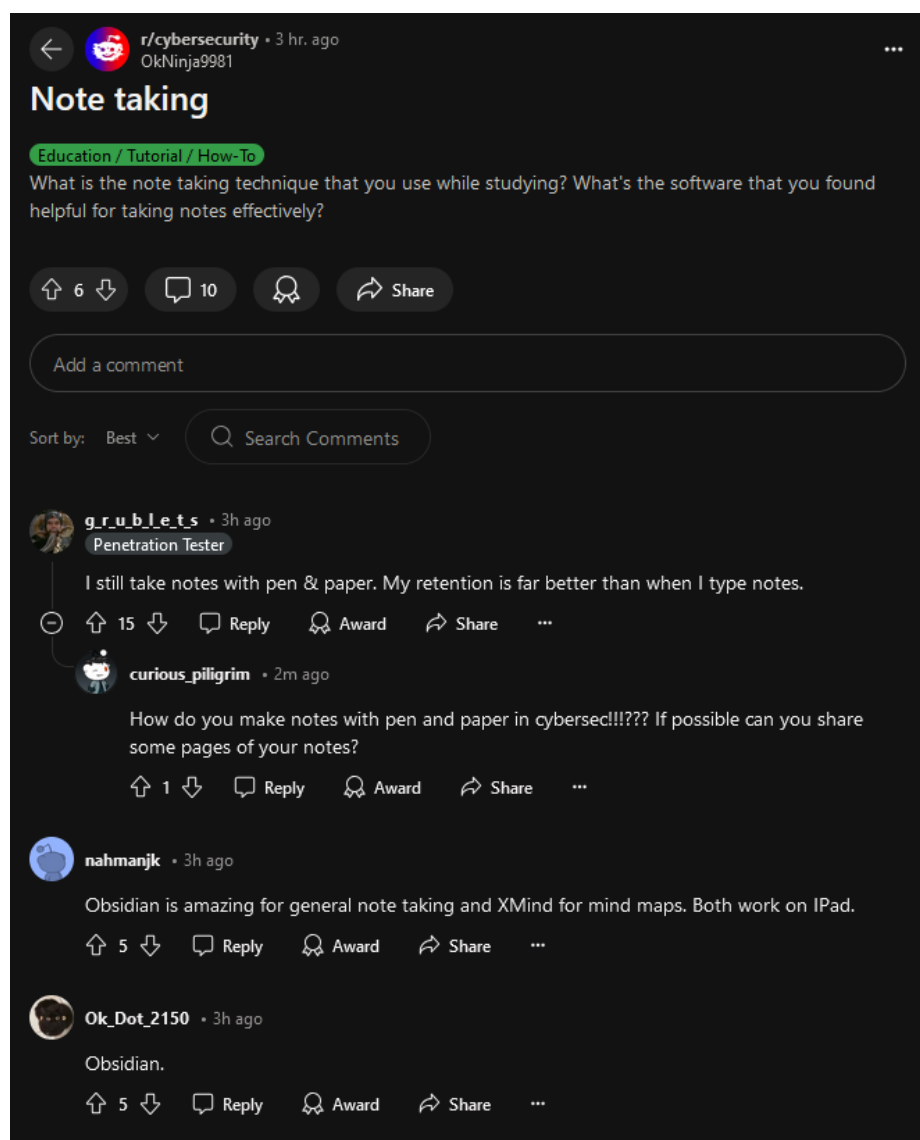


Figure 3.3: This is a screenshot of the comments on a post in the **r/cybersecurity** subreddit. While some posts can be media like a video or image, this post is text, with a question asking other cybersecurity professionals about what kind of note taking techniques they use. Under the text post, members of the community reply with comments with their answer. Each comment may also receive an upvote, additional replies (i.e., a “comment tree”), and potentially “Awards,” which can be purchased to highlight someone’s post or comment and make it stand out in the community.

given a “flair” (a tag) on the post that designates that it was posted by a moderator and meant to discuss something about the community. I give an example in Figure 3.6, where the moderators of **cybersecurity** discuss a rule change in their community, and in many cases members can communicate feedback in the post (like in Figure 3.3, often facilitating discussions about potential rule changes and problems in their communities.

3.2.2 *Empirical Work*

Reddit has been a site for research in social computing scholarship at least ten years since the time of this writing. Reddit has gained a lot of attention as scholars have been interested in how information is created and shared across subreddits (Gilbert, 2013; Leavitt and Clark, 2014), how people come together to form communities (Barbosa et al., 2016), how volunteer leaders manage these communities (Grimmelmann, 2017; Matias, 2019), and the technologies that are used in moderating content on subreddits (Jhaver et al., 2019b). Reddit provides ample opportunities for researchers to study dynamics of communities because of how many subreddit communities exist (at least hundreds of thousands of active communities) and because access to its API was publicly available and relatively unrestricted for most of Reddit’s history, allowing researchers to analyze digital trace data of users and communities to understand or explain different social phenomena. In addition to these benefits, Reddit is also a useful setting for studying variation in community governance because communities are run by volunteers with different backgrounds and strategies and goals for how to manage their community spaces.

Prior research on subreddit communities has investigated the effects that norms have on community success (Chancellor et al., 2018) and how widespread norms diffuse across all subreddit communities (Chandrasekharan et al., 2018). Research by Massanari (2017) evaluated the degree to which Reddit’s macro level norms and design often facilitate toxic “technoculture” of masculinity, making it difficult for marginalized groups to have a voice in certain subreddit’s discussions. Other work has shown how subreddits that center that kind of toxic behavior may spill over into conflicts with other subreddit communities (Kumar et al., 2018). On the other hand, work by Matias (2016b) showed how Reddit can be a site

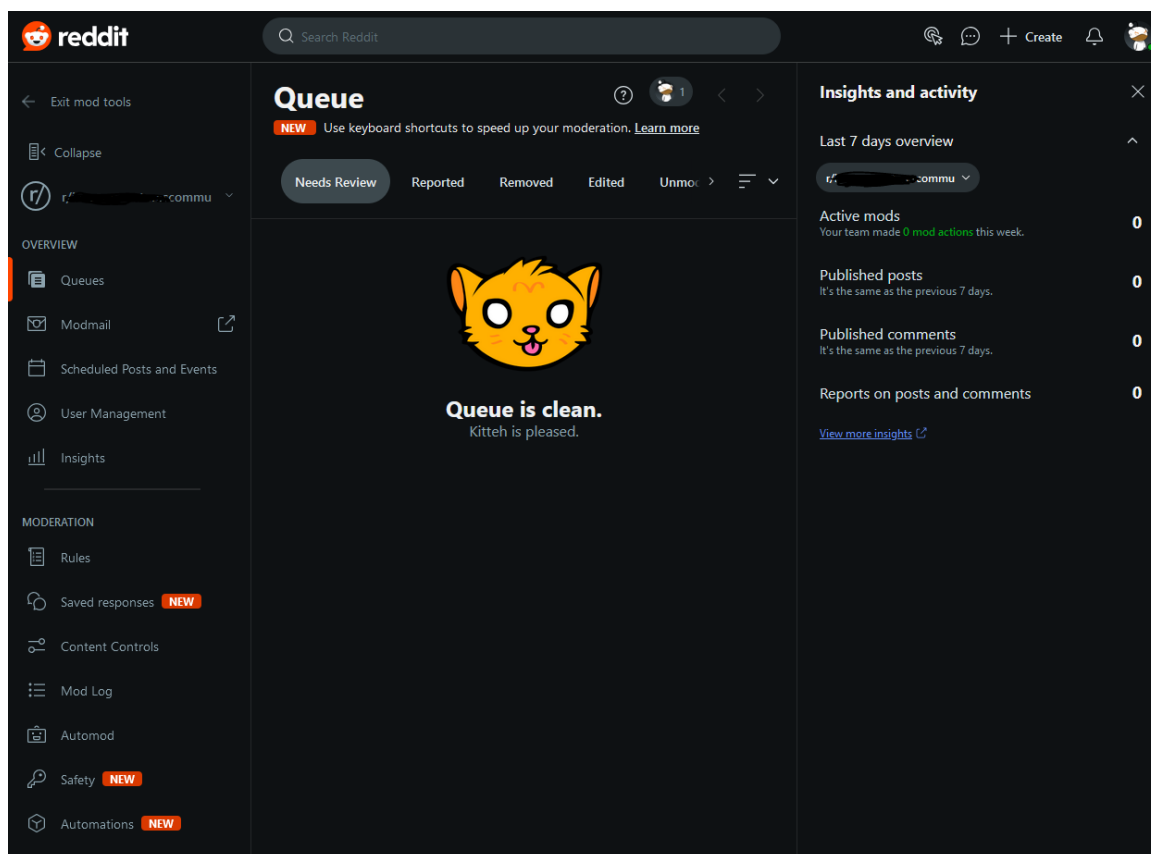


Figure 3.4: This is a screenshot of the user interface for Reddit’s “Mod Tools.” Leaders of subreddit communities access this interface to manage and govern the communities in which they moderate. The left side of the interface shows the subreddit communities the user moderates as well as multiple tools that use to facilitate community governance. For example, “Queues” shows the moderator which posts or comments have been reported / flagged for review by other users. Under that is “Modmail”, which is a messaging system for each subreddit to centralize moderator communication to members. Under the “Moderation” category is a page to create a subreddit’s rules, an option to customize what content types are allowed in the subreddit, a log of all moderator actions in their community, and the “Automod”, which is an automated governance tool that works as a bot for automatically sanctioning users and warning them about rule violations.

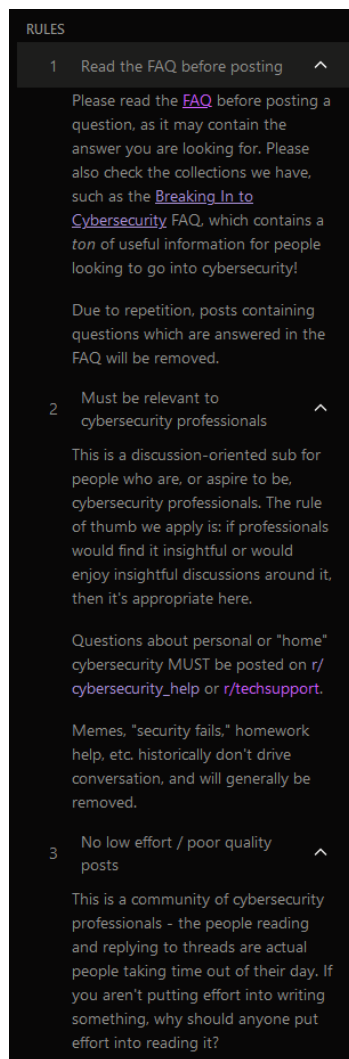


Figure 3.5: This is a screenshot of the rules for the r/cybersecurity subreddit. This is not the full list, but this is what users see on the right side of the interface when browsing the r/cybersecurity subreddit feed.

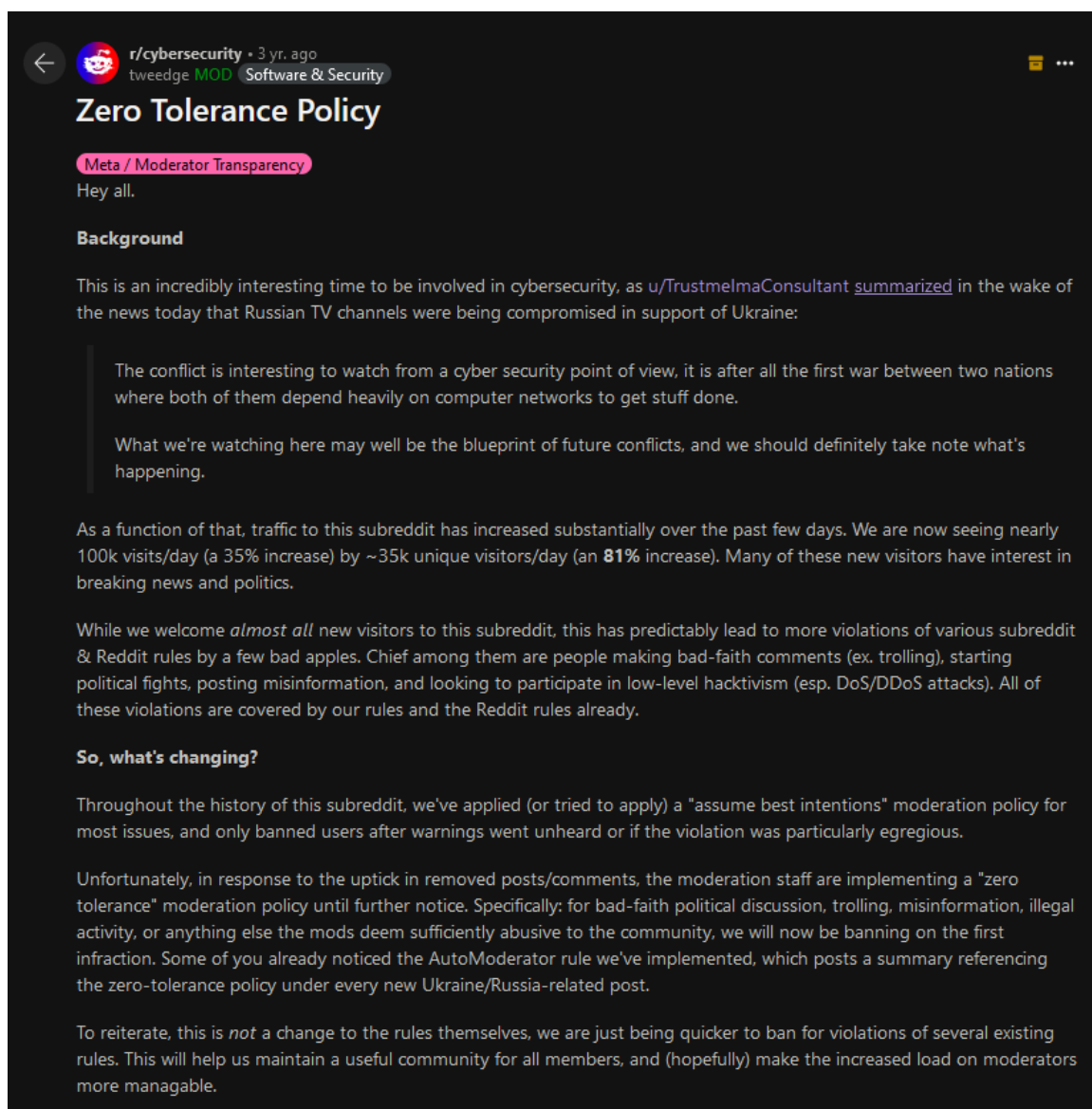


Figure 3.6: This is a screenshot of a post about a Zero Tolerance Policy rule change in the community made to the r/cybersecurity subreddit by one of it's head moderators. The post describes the problem and what is going to change

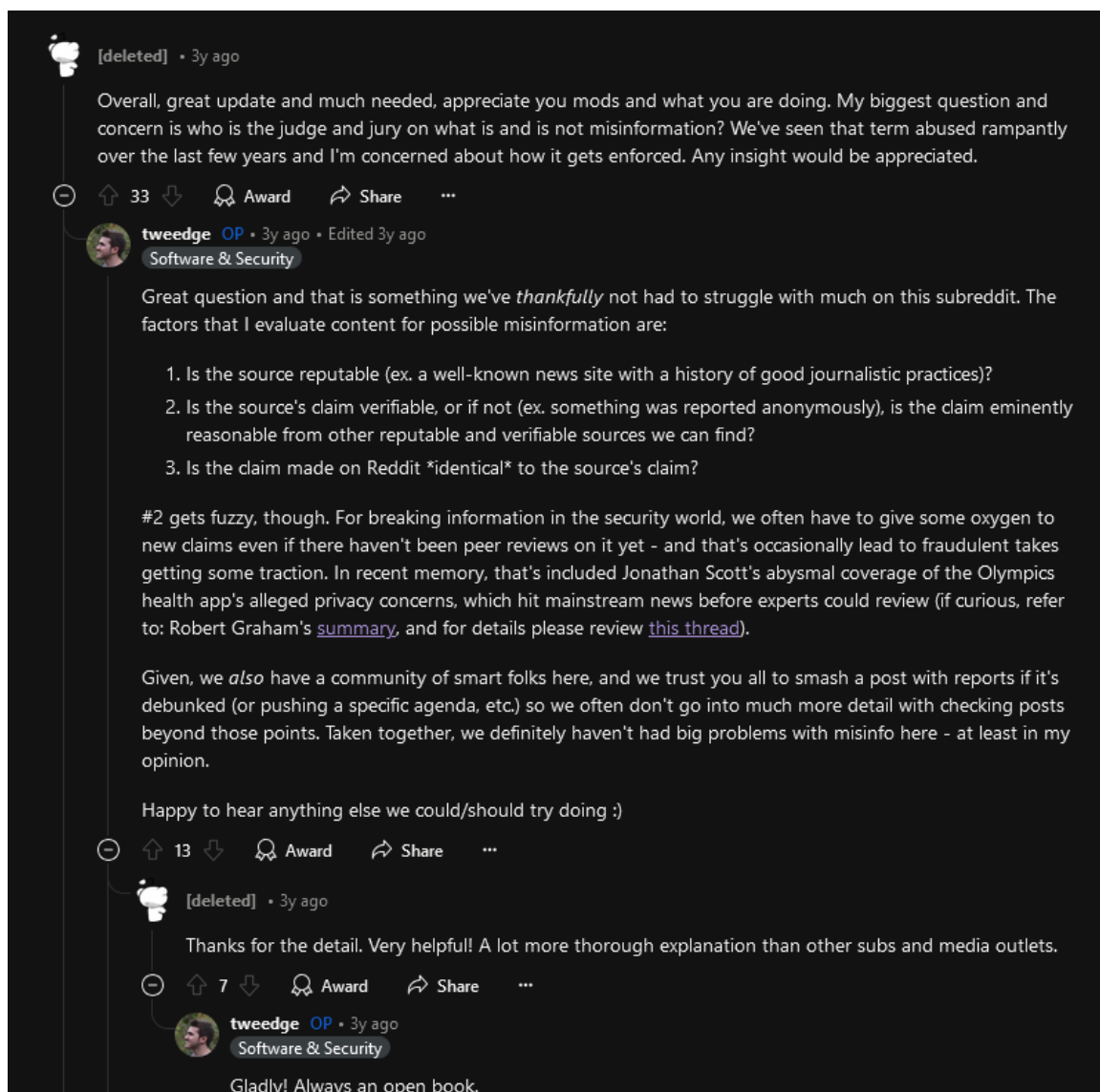


Figure 3.7: This screenshot captures the discussion about the Zero Tolerance Policy rule change by members with the leaders of the `r/cybersecurity` subreddit. Members can ask questions and give feedback in the comments.

for collective action and protest.

Redditors can belong to multiple subreddit communities, and this dynamic has catalyzed research about how this overlap affects the ecology of Reddit’s online communities. Tan and Lee (2015) found that multiple communities with the same topic may emerge, and users will explore these new communities, possibly contributing in both or return to their old community. Work by Hwang and Foote (2021) interviewed members of “small” subreddit communities to understand why they participate there and not in larger communities. This work found that smaller subreddit communities may actually fill a niche that doesn’t require or even want to grow, offering benefits that larger communities can’t fulfill. Similar research by TeBlunthuis et al. (2022) interviewed members of overlapping communities in research that supported Hwang and Foote (2021) but also showed how these smaller communities may serve a role in the ecosystem of an online community topic, with tradeoffs for community characteristics like audience size, homophily, and expertise. Further work by TeBlunthuis and Hill (2022) found that subreddit communities that overlap in topic and members may be operating under a mutualistic rather than competitive framework.

Other research has explored the ways in which subreddit communities evolve and change in response to internal or external problems. Work by Kiene et al. (2016) and Lin et al. (2017) investigated the strategies and outcomes that subreddit communities deployed and experience to manage a massive influx of newcomers. Work by Tan (2018) developed a genealogy graph of subreddit communities to explore community creation over time and how they grow in relation to their parent communities. Finally, more recent work by Reddy and Chandrasekharan (2023) analyzed rule text language using the Internet Archive to evaluate how rules changed on subreddits over time.

Special attention has been brought to the volunteer moderates who govern subreddit communities in social computing research. One study by Matias (2019) interviewed moderators of subreddit communities and described the work they as “civic labor” because they essentially govern public spheres of discourse on the Internet. Work by Seering et al. (2019) interviewed moderators of subreddits (as well as Facebook Groups and Twitch communities) to understand how individuals are recruited, trained, and experience the work of a volunteer moderator of an online community. Other scholars have focused on the psychological

aspect of being a volunteer moderator of subreddit communities. For example, Dosono and Semaan (2019) explored the ways in which subreddit moderators for Asian American Pacific Islander communities practiced emotional labor to sustain their communities. Work by Schöpke-Gonzalez et al. (2022) used surveys to understand why volunteer moderators of subreddits become “burned out” by their work and eventually quit, finding that a lot of it is due to the psychological distress of content they moderate as well as having little time to actually work on their communities. More recent survey work by Bulat et al. (2024) studied the psychology of volunteer moderators on Reddit in the context of how they view democratic governance in their communities, perceptions of fairness, feelings of belonging, and community success outcomes. Finally, extensive work by Li et al. (2022b) used digital trace data of moderation actions of subreddits to estimate the amount of monetary value these volunteers bring to Reddit, the platform that hosts the subreddit communities and is successful only because volunteers moderate the content within their subreddits.

Another line of research in this area focuses on the other side of governance: the subreddit members who are affected by governance actions of moderators and their tools for automating some of those actions. For example, work by Shagun Jhaver sought to understand how users react to having their content removed for violating a subreddit rule (Jhaver et al., 2019a), whether transparency in moderation actions actually matter in reducing future rule violations (Jhaver et al., 2019c), and the benefits and tradeoffs of relying on the automated moderation tool “Automod” in facilitating governance work on subreddits (Jhaver et al., 2019b). Other research has explored the efficacy of sanctioning content on subreddits (Srinivasan et al., 2019) and sanctioning entire subreddit communities for breaking sitewide rules (Chandrasekharan et al., 2017). Scholars have also been interested in the extent to which volunteer moderators of subreddits are transparent at all about their governance actions (Juneja et al., 2020) and how much labor is involved when volunteer moderators have to deal with rule violators who want to contest sanctions (Atreja et al., 2024).

Very few studies have actually studied governance of subreddits in terms of organizational structure and decision-making. Most work related to subreddit governance focuses on the rules. Fiesler et al. (2018) created a broad category of rule types across subreddit com-

munities and analyzed their occurrences. Mentioned earlier, Reddy and Chandrasekharan’s (2023) work on the evolution of rules in subreddit communities is a step to trying to further understand how these thousands of volunteer groups decide on how and when to change their rules. They found that most rule additions occur in less than a year after subreddit creation, and these rule additions are mostly related to behavioral problems that emerged in their communities rather than civility and relevance problems. This study contributes to this body of work by describing the decision making processes in action, how these volunteer leaders think about their rules and why and when they should change them.

3.3 Fandom

Fandom is a website that allows its users to create their own “wiki”, which is a kind of collaborative encyclopedia for all things related to the wiki’s topic. Users can contribute to a Fandom wiki by creating article pages, editing article pages, and helping maintain the pages over time to ensure that updates and new articles are relevant and of high quality. While wikis on Fandom encourage open collaboration, hierarchies are embedded into the social organization of its users. Fandom wikis may have admins, bureaucrats, moderators, and other classes that give such members more authority over the shared space than base user contributors. Such authority also influences rule creation and the ability to edit rule pages.

Users of Fandom may find communities using the search feature on the Fandom home page (shown in Figure 3.8). Users can create their own wiki or join an existing wiki by contributing to its articles and discussions. Unlike subreddits, Fandom wikis are typically organized around the peer production of information artifacts, like the lore of a comic book universe as illustrated in the screenshot of Figure 3.9. Contributors to a Fandom wiki produce articles, like those found in Wikipedia, but these are specifically related to the Fandom’s topic of interest, like the comic book characters of the Marvel universe. I give an example of a Fandom wiki article in Figure 3.10. Community member interactions occur primarily on articles’ “Talk Pages”, which is a space on the article page where editors can discuss improvements or problems with the article. I give an example of these interactions in Figure 3.11, where two members of the Marvel page discuss a potential rule that was

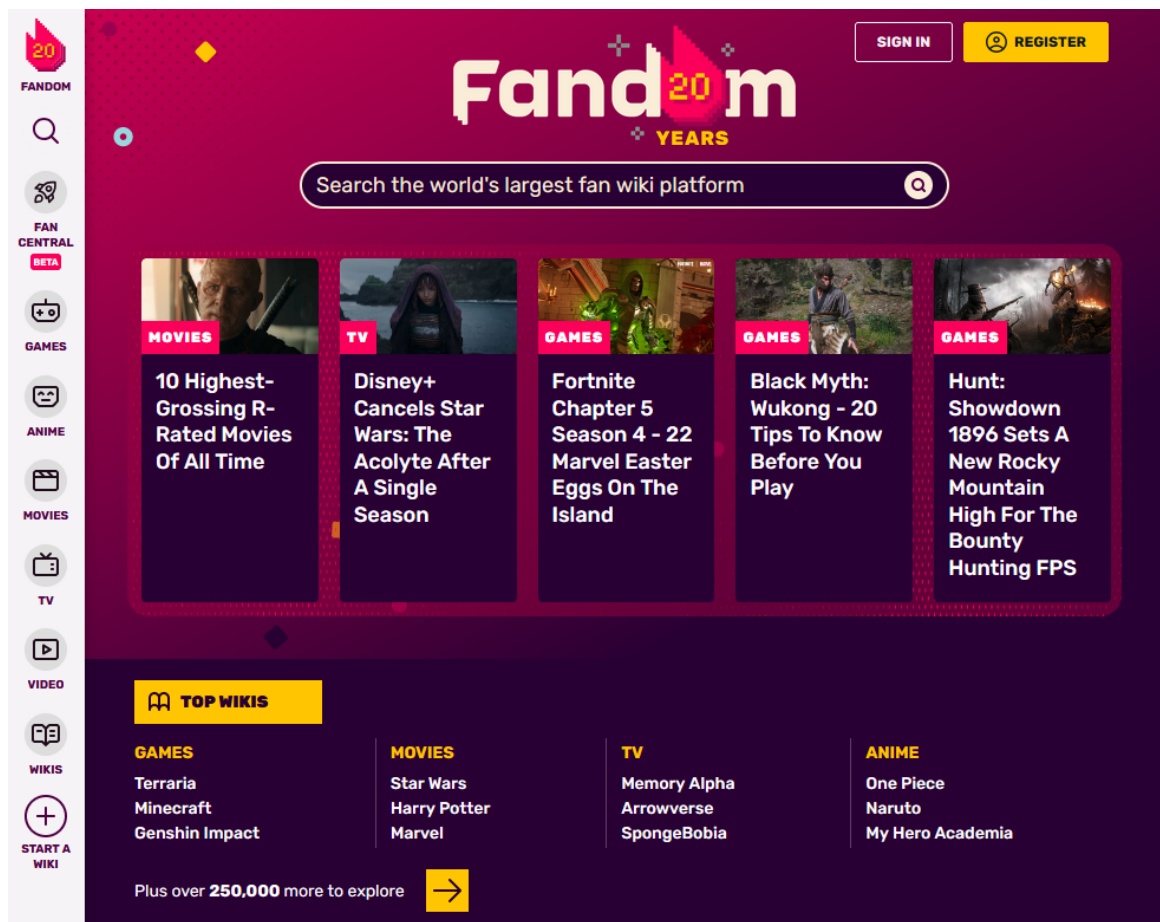


Figure 3.8: This is a screenshot of the main home page for Fandom. Users can search for fan wikis in the search bar or explore suggested wikis.

broken for the article on which they're editing. Some Fandom wikis activate a "Discussions" feature for their community, which essentially creates a forum like space for members of the community to more freely discuss and share information related to the wiki's topic of interest, as shown in Figure 3.12 and Figure 3.13.

3.3.1 Fandom Rules and Governance

Creators of Fandom wikis are given a variety of tools to customize their community space, as shown in the "Admin Dashboard" wiki feature in Figure 3.14. Governance of Fandom

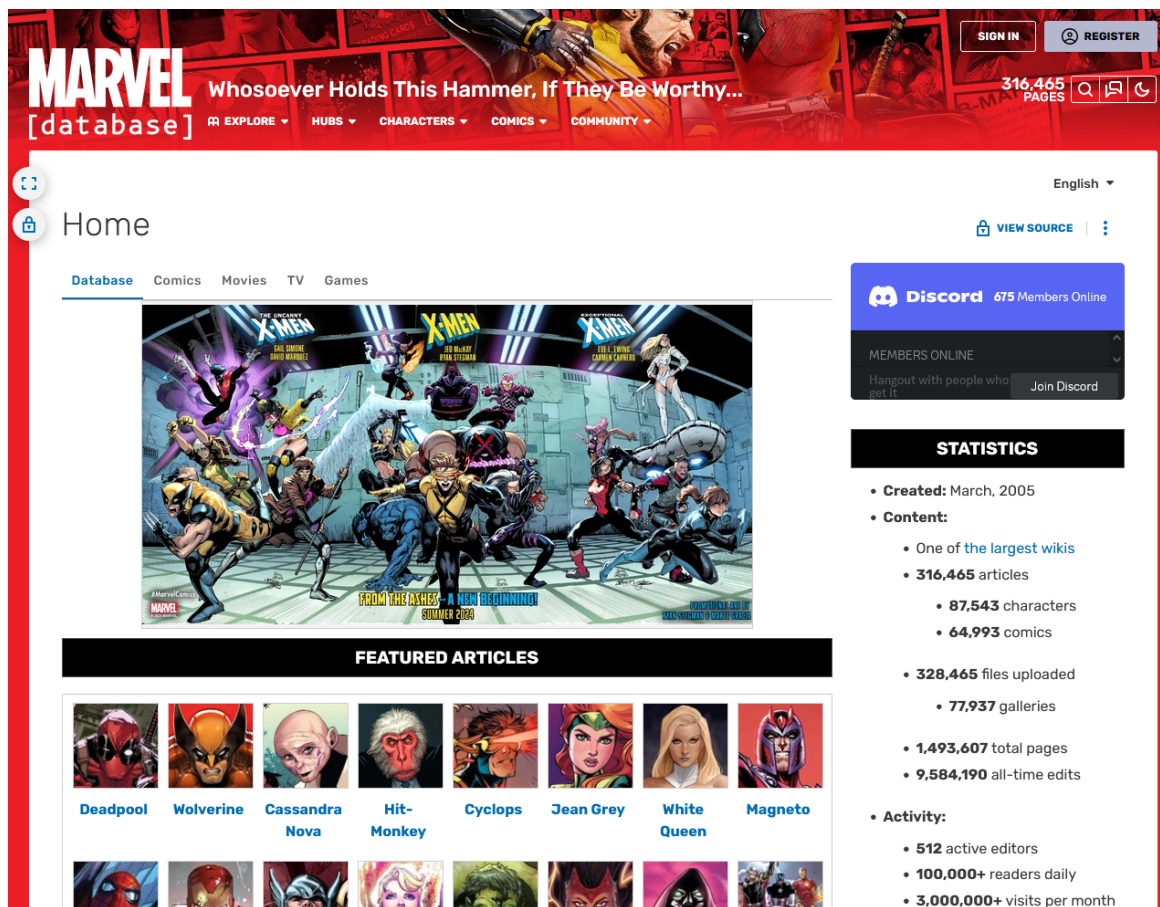


Figure 3.9: This is a screenshot of an example Fandom wiki's, the Marvel Wiki, home page.

in: [Characters](#), [X-Men \(Earth-616\)/Members](#), [X-Force \(Intelligence Agency\) \(Earth-616\)/Members](#), and [182 more](#)
English ▾

James Howlett (Earth-616)

EDIT

“ I'm *Wolverine*. I'm the best there is at what I do, but what I do best isn't very nice. ”

—[Logan](#)^[src]

Cursed with a berserker fury, the violent [mutant](#) known as **Wolverine** has a reputation of both as an outstanding super hero and as a lethal killer. Born as **James Howlett** to a wealthy Canadian family at the end of the 19th century, he was forced to abandon his family after the tragic manifestation of his bestial abilities of accelerated [healing factor](#), keenly enhanced senses and bone claws in each hand. Adopting the name **Logan**, he wandered the world, living a long life filled with blood, war and betrayal,^[89] having vile [Sabretooth](#) as his archenemy,^[90] and the Japanese warrior [Ogun](#) as his sensei.^[91] Unfortunately, by acting as a lone wolf, Logan was an unwitting subject of the [Weapon X Program](#) and had his bones coated in indestructible [Adamantium](#), becoming an even more lethal asset.^[92] Rescued and assisted by [Mac](#) and [Heather Hudson](#), he joined [Department H](#) as a Canadian government operative known as the Wolverine.^{[93][94]}

Later, Logan was invited by [Charles Xavier](#) to join the [X-Men](#), a team of heroes fighting for peace between humans and mutants.^[95] Proving to

Wolverine



Gallery

Name	James Howlett ^[1]
Current Alias	Wolverine
Aliases	[Expand]

Figure 3.10: This is an example of an article page on the Marvel Wiki. Members can edit the article, add new content, revert new edits, and discuss aspects about the article on the article's Talk Page.

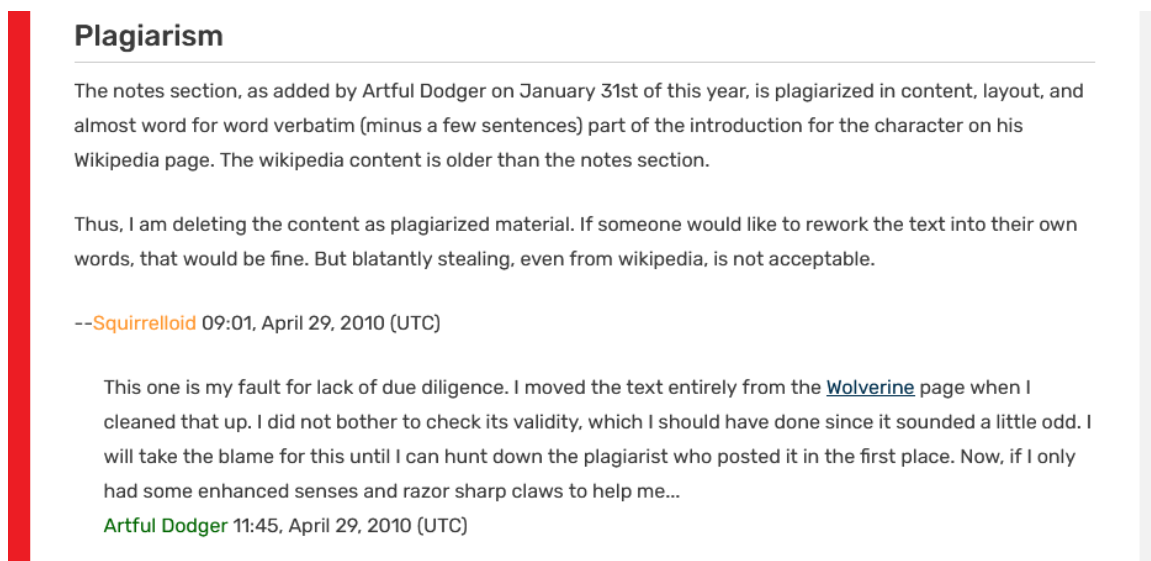


Figure 3.11: Here is a screenshot of a “Talk Page” for the Wolverine article on the Marvel fandom wiki. In this example of a Talk Page, members are discussing possible plagiarism of text in the Wolverine article page.

wikis is left up to community leaders, although a newly created wiki is given a default set of rules (see Figure 3.15). The creator of a Fandom wiki can determine what “rights” each user has as well as create user groups that correspond to the amount of permissions that user has in editing the wiki and managing its members. For example, in Figure 3.16, the Marvel Fandom wiki has organized its internal leadership into a rather strict hierarchy starting with “Bureaucrats”, who are the head administrators, at the top, followed by regular “Administrators” who just help with regulating policy issues and newcomer questions, which is then followed by “Moderators”, who are tasked with moderating content on the wiki articles or in the Discussions forums. Former administrators are recognized by this Fandom wiki’s leadership, and even “Bots”, automated tools to help protect wikis against vandalism or automate editing processes, are understood as part of the internal organization of the community’s administrators.

While the default rule set for wikis has five generic rules about participating, more established wikis like the ones I studied in this project tend to add more and complex rules.

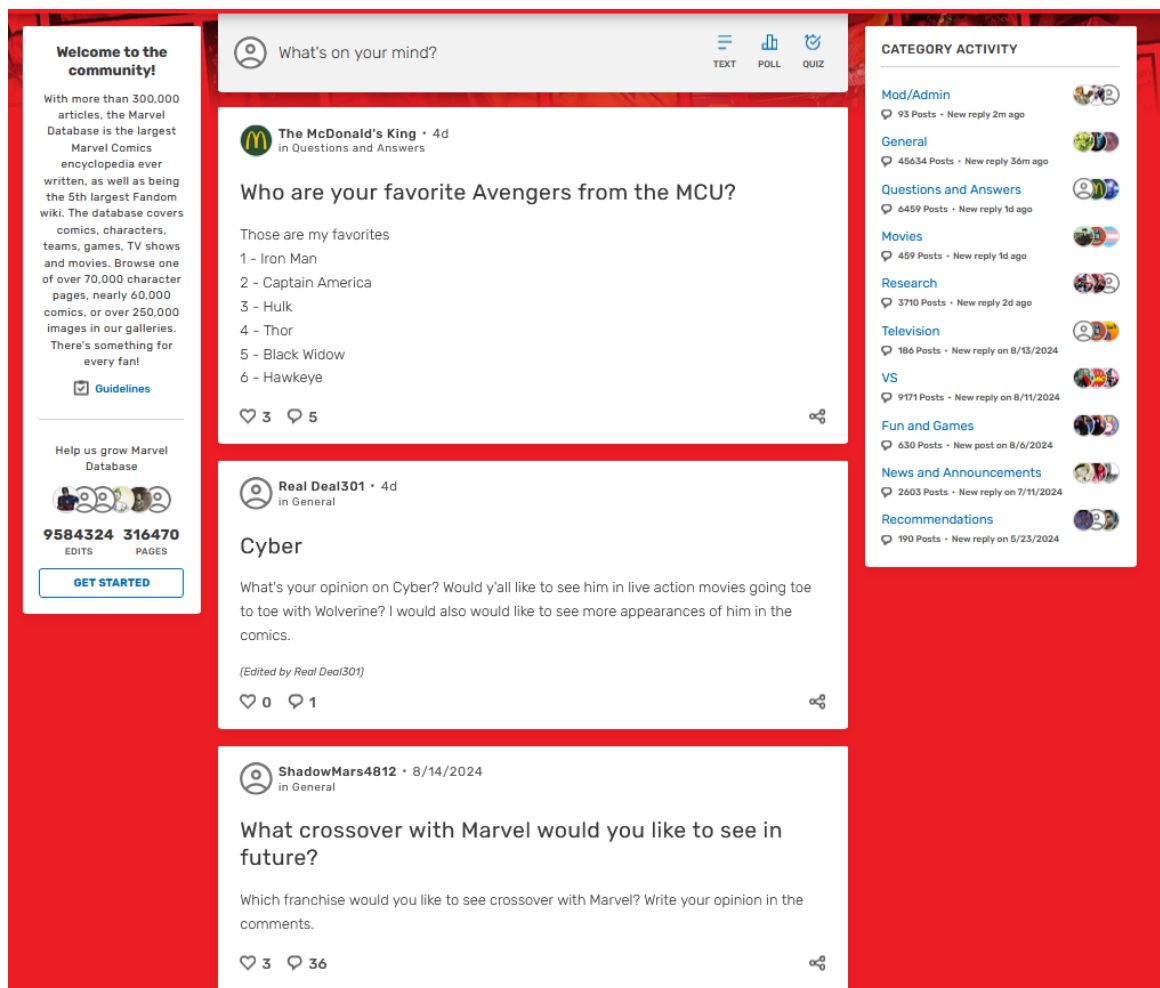


Figure 3.12: This is a screenshot of the “Discussions” page in the Marvel fandom wiki. Discussions functions similarly to any other online forum, with users able to add comments and rate posts with “likes.” Not every wiki has a Discussions forum, as it’s an opt-in choice for wiki community leaders.

[← ALL POSTS](#)

ShadowMars4812 • 8/14/2024
in General

What crossover with Marvel would you like to see in future?

Which franchise would you like to see crossover with Marvel? Write your opinion in the comments.

♥ 3 💬 36



[VIEW OLDER REPLIES](#)



HorrorMosco • 5d

Godzilla should be good



RJ Plotner • 5d

[@HorrorMosco](#) Put Godzilla, King Kong, and Mothra in there and it would be golden.



Joshua Yates • 4d

How about a Maximum carnage movie

(Edited by Joshua Yates)

Figure 3.13: Here is an example of a post with comments from members in the Marvel fandom wiki.

Figure 3.17 shows a screenshot of the article page for the Marvel Fandom wiki’s “Policies and Guidelines” (i.e., the rules of the community). Compared to the default rule set, this is much more extensive. Each policy in this list is essentially just a category with more in depth and nuanced information in its corresponding article page. For example, the “Article Guidelines” policy page contains numerous sub-sections (see Figure 3.18 related to formatting and information requirements for contributing an article to the Marvel Fandom wiki). Within each sub-section of the Article Guidelines policy page is more nuanced information about how to correctly contribute to the wiki, as I show in Figure 3.19. When a Fandom wiki has activated the Discussions feature, they may also incorporate rules for participating in the forum. I give an example of those rules for the Marvel Fandom’s Discussions in the screenshot in Figure 3.20. Finally, rule change discussions may occur on the talk page of a Fandom wiki’s rules article. I show an example in Figure 3.21 where members of the Marvel Fandom wiki discussed a rule change related to how the info boxes on articles for comic book characters should be formatted. Members of the community responded with their opinions about the change and cast their vote.

3.3.2 *Empirical Work*

Research on wikis, the underlying infrastructure that supports the type of peer production community for Fandom wikis, has resulted in a large body of work. Wiki communities are their own unique case study as they are centered around the peer production of a public good, the most well known of which is Wikipedia, the peer produced encyclopedia. Wiki communities can come together to produce information on almost any topic, from the lore of comic book universes to the documentation for open source software. While a large majority of this research is on Wikipedia, especially with respect to research on online governance, much of this work has implications for research on Fandom wikis which share the same technical infrastructure and code as Wikipedia and many of the same social conventions and organizing principles as well. That said, I overview the empirical work on wiki communities and governance in this section.

Past research has explored the creation of new wiki communities and how they often seed

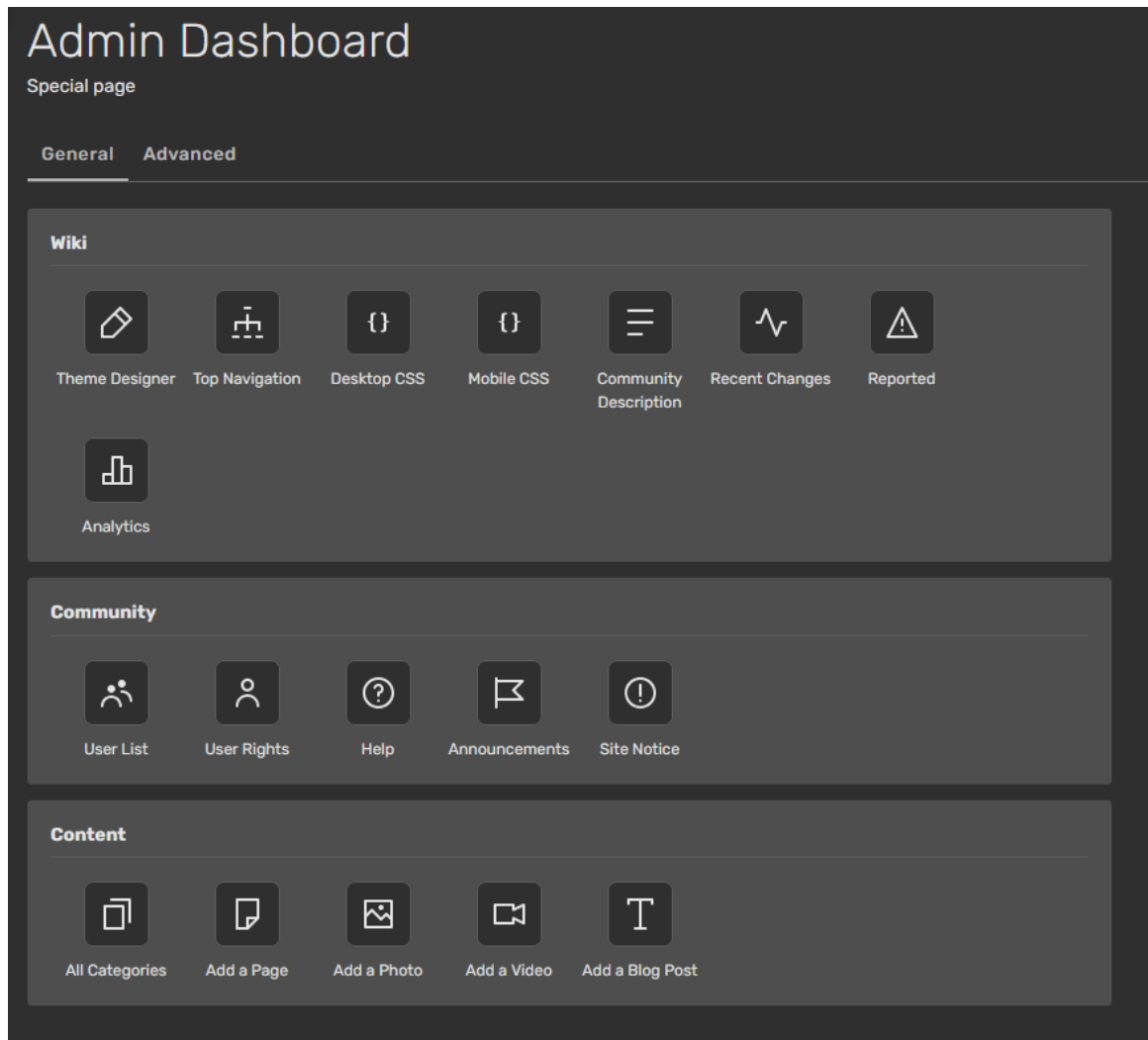


Figure 3.14: This is a screenshot of the “Admin Dashboard” for Fandom wiki leaders. Like the Mod Tools on Reddit, this dashboard gives leaders agency over how their wiki is designed as well as how it manages members and member permissions.

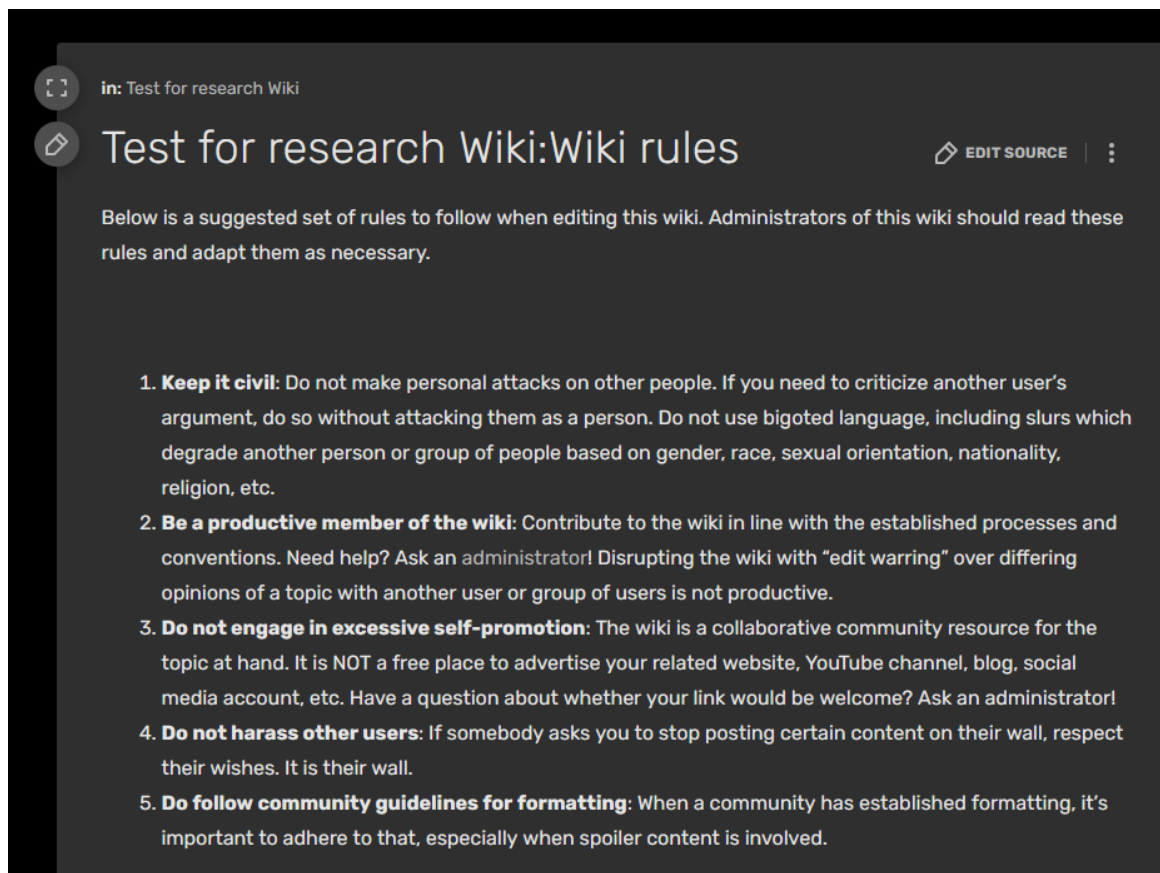


Figure 3.15: This is a screenshot of the default rules that are automatically generated for any newly created wiki on Fandom.

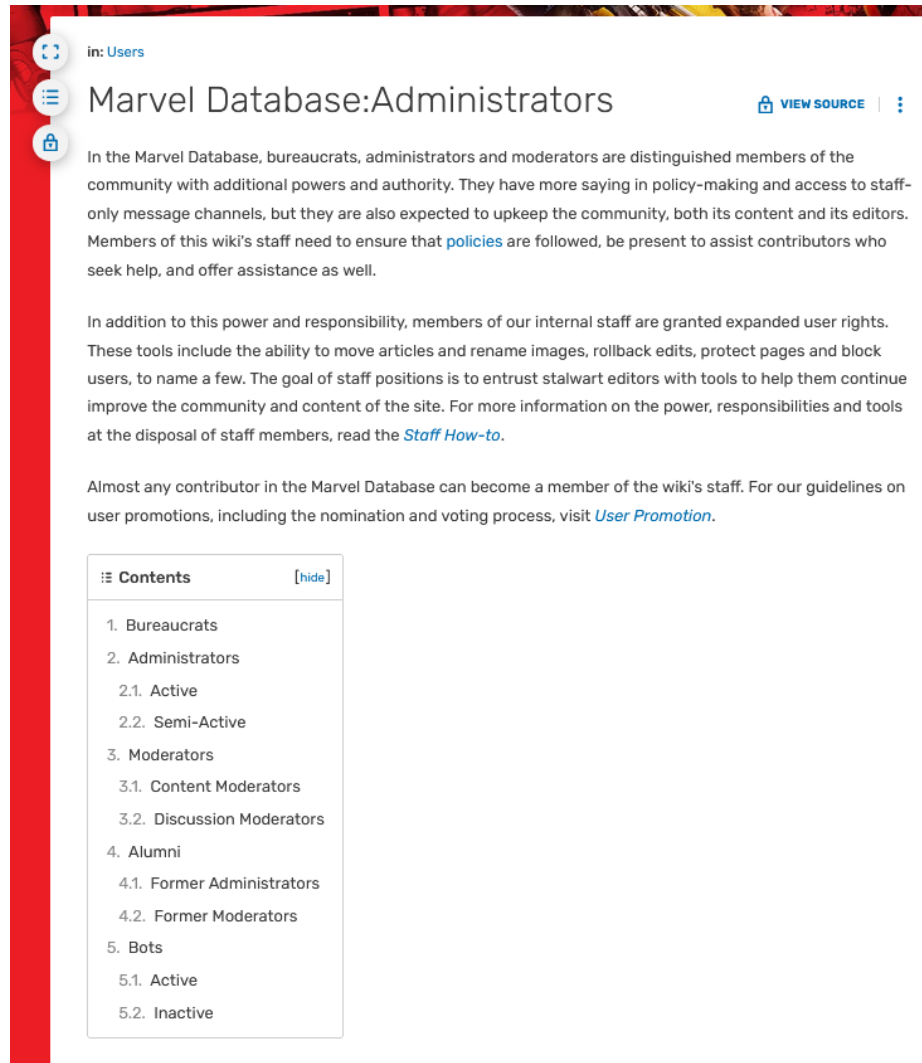


Figure 3.16: This is a screenshot of the Marvel fandom wiki's information page on its leaders, titled "Administrators." On some wikis, like this one, leaders create a hierarchy among themselves. Here, "Bureaucrats" are the head administrators of the Marvel fandom wiki. After them, regular "Administrators" hold power to help manage the wiki and organize the community's "Moderators", who are primarily given enough power to moderate content on article pages or in the discussion forums.

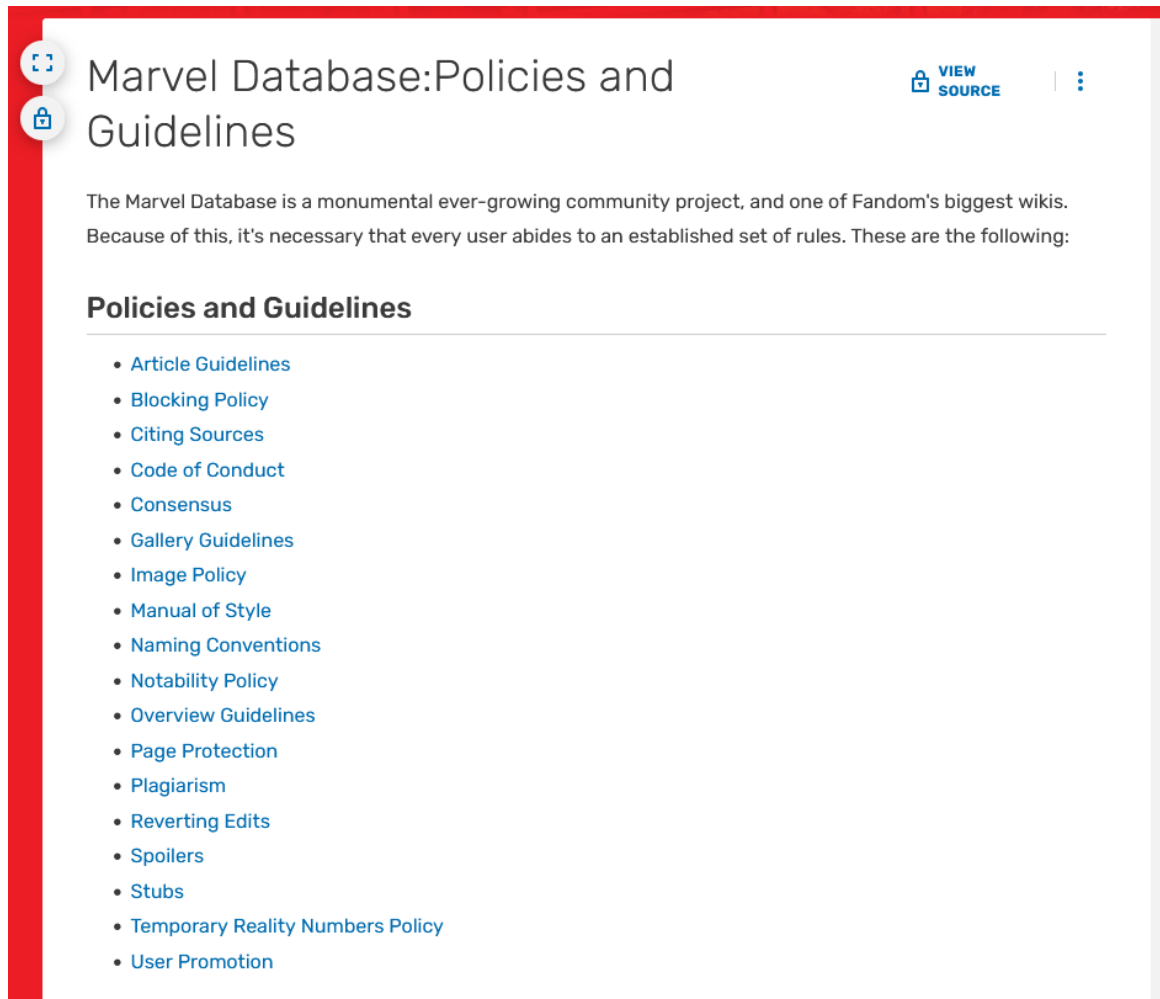


Figure 3.17: This is a screenshot of the list of rule *categories* (“Policies and Guidelines”) of the Marvel fandom wiki.

 Contents	[hide]
1. Article Creation	
2. Article Naming	
3. Article Layout	
3.1. Infobox	
3.2. Overview	
3.3. History	
3.4. Personality	
3.5. Attributes	
3.6. Paraphernalia	
3.7. Appearances	
3.8. Summary/Synopsis/Plot/Story	
3.9. Cast	
3.10. Notes and Trivia	
3.11. See Also	
3.12. Links and References	
4. See Also	

Figure 3.18: A screenshot of the Table of Contents for the first rule category on the Marvel wiki's Policies and Guidelines Page.

Article Naming

Every article in the Marvel Database must be created with a name that adheres to the [Naming Conventions](#), where the full extent of these guidelines are detailed. The following are only basic elements to take into account for the naming of articles:

- Use title case, in which all words are capitalized, with the exception of certain subsets like articles, short prepositions and conjunctions.
- Since Marvel makes use of the concept of [alternate realities](#), the Marvel Database covers many different versions of a given character or group that exist in other [realities](#). Because of this, the articles for characters, teams and organizations should always include between parenthesis the numeric designation of the universe they belong to (e.g. the article for the Avengers is [Avengers \(Earth-616\)](#), "Earth-616" being the numeric designation for the [Prime Marvel Universe](#)).
- The only type of pages pertaining to an individual or group that don't feature a reality designation in their title are disambiguation pages since these articles serve as a hub which lists a character's different counterparts from different universes.
- Article titles for items, vehicles, races and locations don't require reality designations since all versions from different universes of either of these are grouped in the same article.
- Character articles should always be named after the character's real name if possible, avoiding diminutives (e.g. the article for the Iron Man of the Prime Marvel Universe is [Anthony Stark \(Earth-616\)](#), not "Tony Stark (Earth-616)")
- Comic volumes should be named based on the title displayed in the [indicia](#) of their respective issues.
- Avoid including "The" in the title of articles.

Figure 3.19: A screenshot of the first policy in the Article Guidelines of the Marvel fandom wiki.

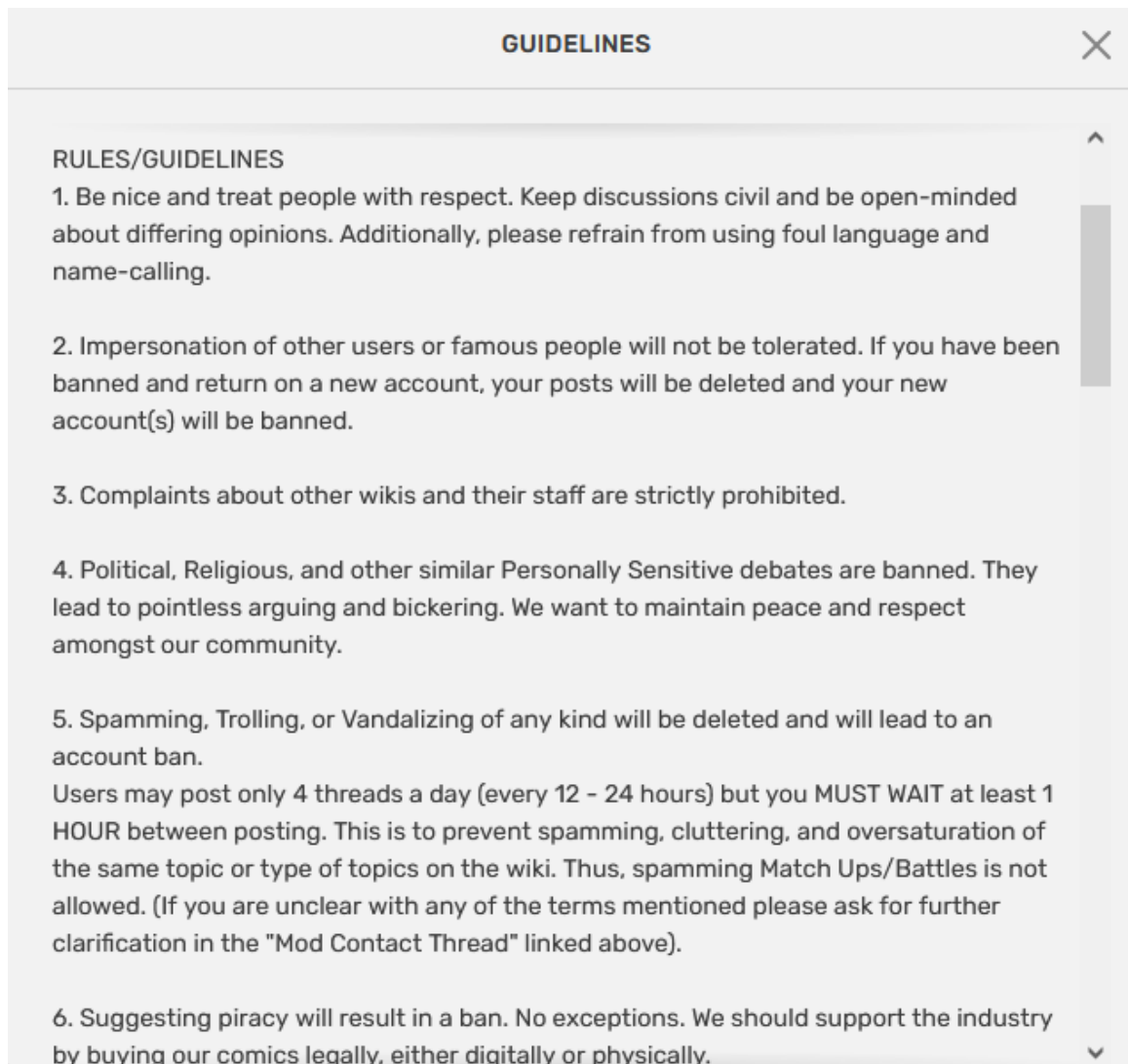


Figure 3.20: A screenshot of the rules of the Marvel fandom wiki's Discussions rule set. Note that rules here are more related to interacting with other people and not rules related to formatting article pages.

Talk: Article Guidelines SIGN IN TO EDIT

[← Back to page](#)

Character infobox title

We had a discussion on [Thor's talk page](#) about the correct use of infobox title field. Two preferred ways were using a **common** or **current** name and I think that it would be beneficial to choose between those. You could of course vote for other options also. Please indent your replies with an asterisk and bold your preferred way for helping to keep track of a discussion. —[Mrkermit](#) (talk) 10:11, May 21, 2016 (UTC)

EDIT:In the case I was little unprecise with options, I want to clarify that *common name* means best-known alias or real name and *current name* means current alias. In our initial discussion those were *Thor* and *Odinson*. —[Mrkermit](#) (talk) 16:44, May 21, 2016 (UTC)

- **Common** names are more recognizable for readers. —[Mrkermit](#) (talk) 10:11, May 21, 2016 (UTC)
- Don't feel the existing system of **Current** need be changed. -- [Annabell](#) (talk) 10:31, May 21, 2016 (UTC)
- I prefer using the **current** system. Using the common names would be inaccurate as there's characters who change alias all the time, like [Hank Pym](#). Listing him as still being Ant-Man (or whatever is more recognizable for readers) would be totally inaccurate and would also confuse readers.
 - [The Many-Angled One](#) (talk) 15:23, May 21, 2016 (UTC)
 - Poor Hank has had so many aliases that his common title can only be *Hank Pym*. —[Mrkermit](#) (talk) 16:44, May 21, 2016 (UTC)
- **Current**. Common names are what disambigs are for. --[Spencerz](#) (talk) 15:30, May 21, 2016 (UTC)
 - Sorry but I don't understand, we have disambiguation pages for every kind of names. Besides, I don't see how those are connected to infobox titles. —[Mrkermit](#) (talk) 19:13, May 25, 2016 (UTC)
- I don't have much to add since others already said what I think, but yeah I say keep the **Current** system. ([SunGodKizaru](#) (talk) 15:48, May 21, 2016 (UTC))
- I choose **current** as well.--[The ADour-incible ADour](#) (talk) 16:38, May 21, 2016 (UTC)

Figure 3.21: This is a screenshot of the Talk Page for the Article Guidelines policy section on the Marvel fandom wiki. In this example, members are proposing a rule change about how to format an “info box” on an article, with members discussing whether or not they would be okay with such a change.

content and articles from third party sources (Solomon and Wash, 2012). Similar work has investigated the motivations and goals of founders of wiki communities, finding a diverse set of reasons while seek to occupy narrow topics and don't seek a large group of contributors (Foote et al., 2017). Other research has focused on problems of survivability of wiki communities and membership overlap, finding that wikis often benefit when they share members who are longstanding members of other wiki communities (Zhu et al., 2014). A frequently researched problem in the context of wikis is retaining newcomers and managing growth. Work by Ransbotham and Kane (2011) investigated membership turnover in Wikipedia, finding that it might be experienced more than is ideal. Halfaker et al. (2011) also found that newcomers to Wikipedia may be affected by strong governance actions like revisions (i.e., when an article is reverted to its previous state after a user tries to edit it) and less likely to stay as a contributor.

While some scholarship on wikis is interested in the quality of contributions (Arnold et al., 2017), I focus specifically on the more widespread problem of how governance is negotiated and enacted in wikis. Past research, for example, has explored how wikis enforce policies and norms through banning, article deletions, and community member vigilance (Geiger and Ribes, 2010; Geiger and Ford, 2011; Kayhan, 2015). Other work has focused on how quality control systems like removing poor quality contributions to wikis or requiring accounts deter growth and retention of new contributors (Halfaker et al., 2013; TeBlunthuis et al., 2018; Hill and Shaw, 2021).

More relevant to the questions motivating this research are the works of Butler et al. (2008), Shaw and Hill (2014), and Keegan and Fiesler (2017). Butler et al.'s (2008) research on Wikipedia investigated how a community built on an egalitarian and open collaboration system may turn into hierarchical bureaucracies to handle their growth and popularity. They found that wiki communities may evolve through the accumulation of rules and policies and roles of authority to make changes to the wiki more structured and preserve quality. They also discovered that rules may serve different functions, moving beyond organizing actions to constructing identity of a community and signaling to external stakeholders. Shaw and Hill's (2014) work on the evolution of wiki leadership used data from Fandom (although it was branded as "Wikia" at the time) and showed that despite beginning as more open and

egalitarian communities, wikis that grow may over time consolidate authority to a small handful of administrators and resemble more oligarchical forms of governance. Finally, Keegan and Fiesler (2017) work on the evolution of rules on Wikipedia revealed an environment in which rule-making is an ongoing process where members frequently deliberate on the state of the rules and how to change them. In sum, the past work on wikis suggests that governance can vary depending on factors like community size and age, and the ways in which authority is distributed among members, often shifting from open, collaborative structures to more hierarchical systems as the community grows and matures.

3.4 Fediverse

The Fediverse functions differently from the previous two platforms in that it's not a central platform at all. Instead, communities on the Fediverse are self-hosted by individuals on various different servers and other pieces of technological infrastructure who then connect to others through the ActivityPub protocol. The most common platforms for microblogging in the Fediverse are Mastodon, Pleroma, Akkoma, and Misskey. Governance is enacted by individual instance owners or administrators. A part of the appeal for many instance administrators and users is having agency over their social network, choosing who they connect to and the kind of content they want to see. On the Fediverse, people share content, engage in discussions, and connect with others in a similar capacity to Twitter. On Fediverse platforms like Mastodon, the act of sharing content in a microblog post is common referred to as a "toot." The interface is quite similar, with a "timeline" (or "feed") showing posts from people the user follows, comments or how many times the post has been shared with other users, as well as search / exploration features to find new people or topics to follow. An example of the interface for one instance, `sfba.social` is shown in Figure 3.22, and the way that people interact with each other on Fediverse posts is shown in 3.23.

To become a member of a Fediverse instance, a user has to first sign up as if they are creating an account on a new platform. Instance administrators may keep this registration process "open" to bring in new members or "closed" to prevent any new members from joining. They may additionally require new members to be vetted in the application process. In the screenshot of the registration process for the `sfba.social` instance in Figure 3.24,

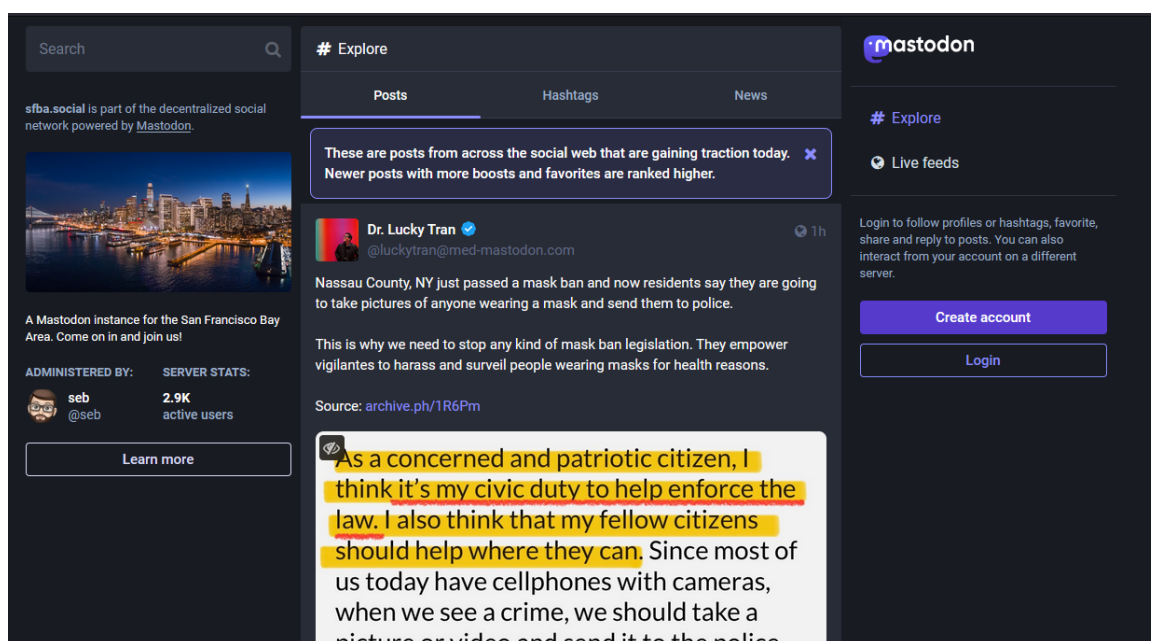


Figure 3.22: This is the interface for a Fediverse instance, `sfba.social`.

the new member would have to agree to the “ground rules” of the instance before they can finish signing up, and in some instances, new applicants may be denied if the administrators believe the applicant isn’t there in good faith (i.e., to spam, advertise, or troll its members). Almost every instance has an “About” page describing its goals, who the instance is for, and potentially a list of its rules. I show a screenshot of the `sfba.social` about page in Figure 3.25.

3.4.1 Fediverse Rules and Governance

Governance in Fediverse instances occurs in primarily two ways. First, instance admins can determine who is allowed to have an account on their community, what kind of content is allowed to be posted, and how people on their instance interact with each other. Administrators may have to do some content moderation by enforcing the rules for participating, like those listed for `sfba.social`’s instance in Figure 3.26. These rules are almost always related to behavior and content because issues like formatting are as prevalent in

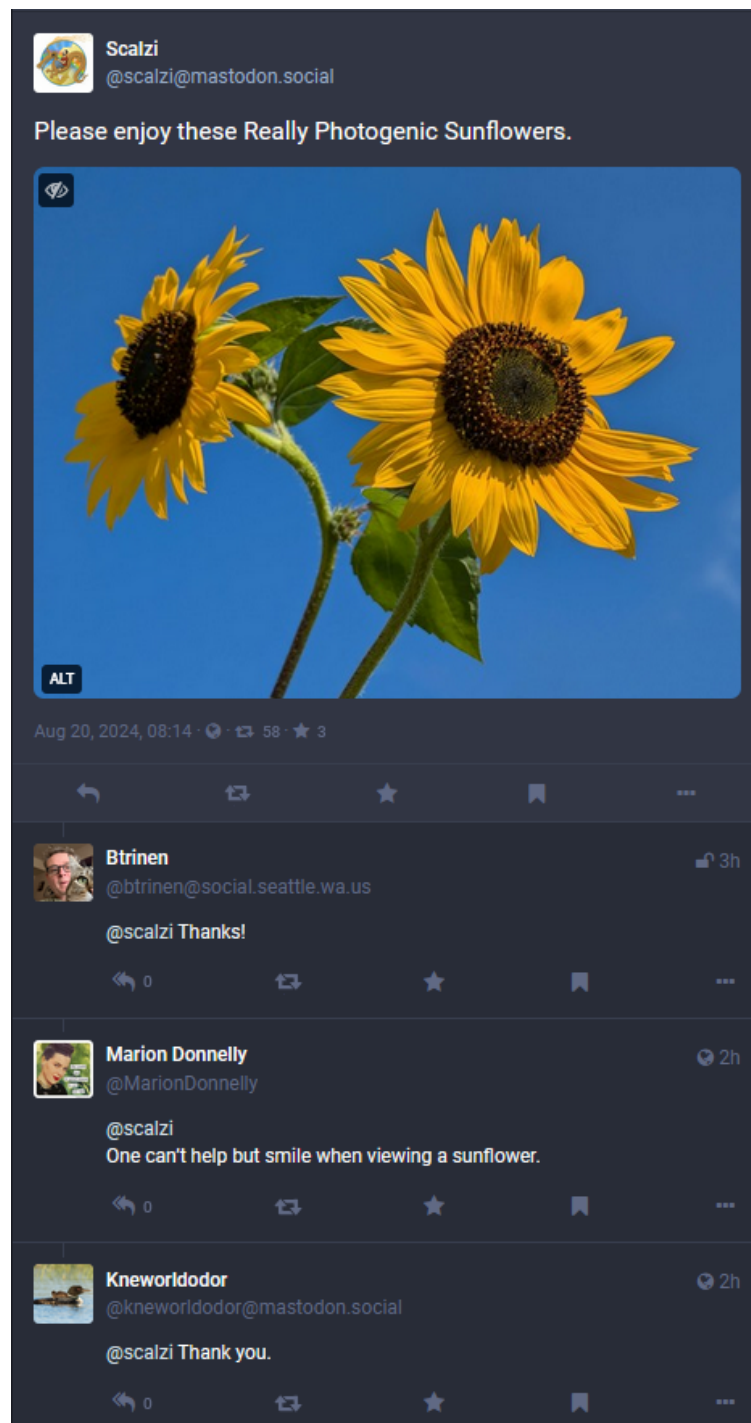


Figure 3.23: This is an example of how users of Fediverse interact with each other. Note that interactions are similar to those on social network sites like Twitter, where people can comment below a post.

The screenshot shows the Mastodon registration interface on a dark background. At the top is the Mastodon logo. Below it is a progress bar with three steps: 'Accept rules' (active), 'Your details', and 'Confirm e-mail'. The main heading is 'Some ground rules.' followed by the text 'These are set and enforced by the sfba.social moderators.' Below this is a list of seven rules, each preceded by a numbered circle. At the bottom are two buttons: 'ACCEPT' and 'BACK'.

mastodon

Accept rules Your details Confirm e-mail

Some ground rules.

These are set and enforced by the sfba.social moderators.

- 1 Do not post illegal, illicit, or unlicensed content.
- 2 Do not use slurs or discriminatory or hateful jokes and language, including (but not limited to) racism, sexism, homophobia, and transphobia; including tone policing or critiquing lived experience.
- 3 Add content warnings to sensitive content (e.x. nudity, violence)
- 4 Do not harass anyone, post other people's personally identifying information ("doxing"), or incite violence against anyone.
- 5 Do not post in bad faith, engage in name-calling, or ad hominem attacks. Criticize ideas, not people.
- 6 Do not post spam (e.x. a post with only a link or an image) or promote crypto-related content such as cryptocurrencies or NFTs.
- 7 Follow the community code of conduct: <https://hub.sfba.social/code-of-conduct> Use content warnings (CW) and #hashtags liberally. Be nice and have fun!

ACCEPT

BACK

Figure 3.24: This is a screenshot of the “registration” process on a Fediverse server, `sfba.social`. Note that potential new members to this community have to read and accept the rules before they can finish signing up. Other registration processes on the Fediverse might be more involved, such as requiring approval from the instance administrator before finalizing account creation.

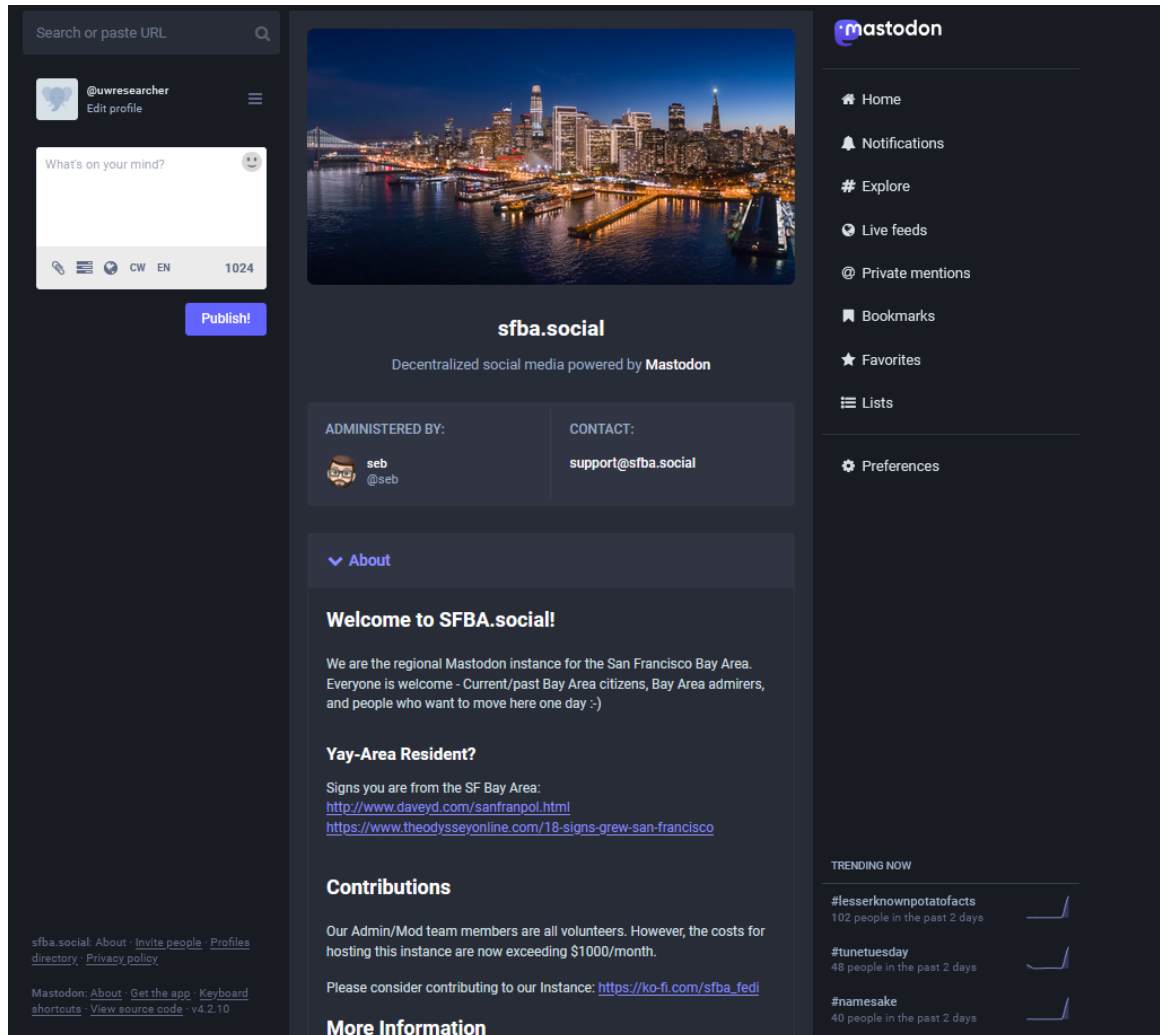


Figure 3.25: This is a screenshot of the “About” page of a Fediverse instance, **sfba.social**.

this kind of micro-blogging setting. The other kind of governance actions that occur for Fediverse instances is the act of “defederating” or “moderating” *other* Fediverse instances. To “defederate” from or “moderate” another instance means that the administrator decides to block interactions with that instance, preventing users on their instance to see or interact with content from the blocked instance. Some instance administrators publicly show the instances from which they’ve defederated, like in the example in Figure 3.27. Defederation may occur because another instance is entirely full of spam and advertising or because the blocked instance has a problem with antisocial behaviors that aren’t addressed, like hate speech or harassment.

3.4.2 *Empirical Work*

Compared to subreddits and wikis, there is comparatively less research on Fediverse communities. This is perhaps a result of the fact the Fediverse didn’t really take on as social networking alternative to more mainstream sites like Twitter and Facebook until 2016, when Mastodon, one of the largest platforms on the Fediverse, was created. Social computing scholarship about the Fediverse is sparse, and governance in the Fediverse has only been studied in mostly the last five years. Nevertheless, I present an overview of the available work on the Fediverse.

Research on the Fediverse has analyzed the novelty of its infrastructure and how the federation protocols that Fediverse instances use allow them to connect to each other as well as the challenges they face in this sociotechnical arrangement (Raman et al., 2019; Zulli et al., 2020). Fediverse platforms look remarkably similar to the interface used by Twitter, and scholars have been especially interested in how the Fediverse has handled migrations of users from Twitter to Fediverse platforms like Mastodon (Dunbar-Hester, 2024). Specifically, some research has found that larger instances have found it challenging to find sufficient resources for moderating their instances with these influxes of newcomers from Twitter (Anaobi et al., 2023).

Research on governance in the Fediverse has focused almost exclusively on the practice of “defederation”, or blocking other instances. Past work by Gehl and Zulli (2023) showed how

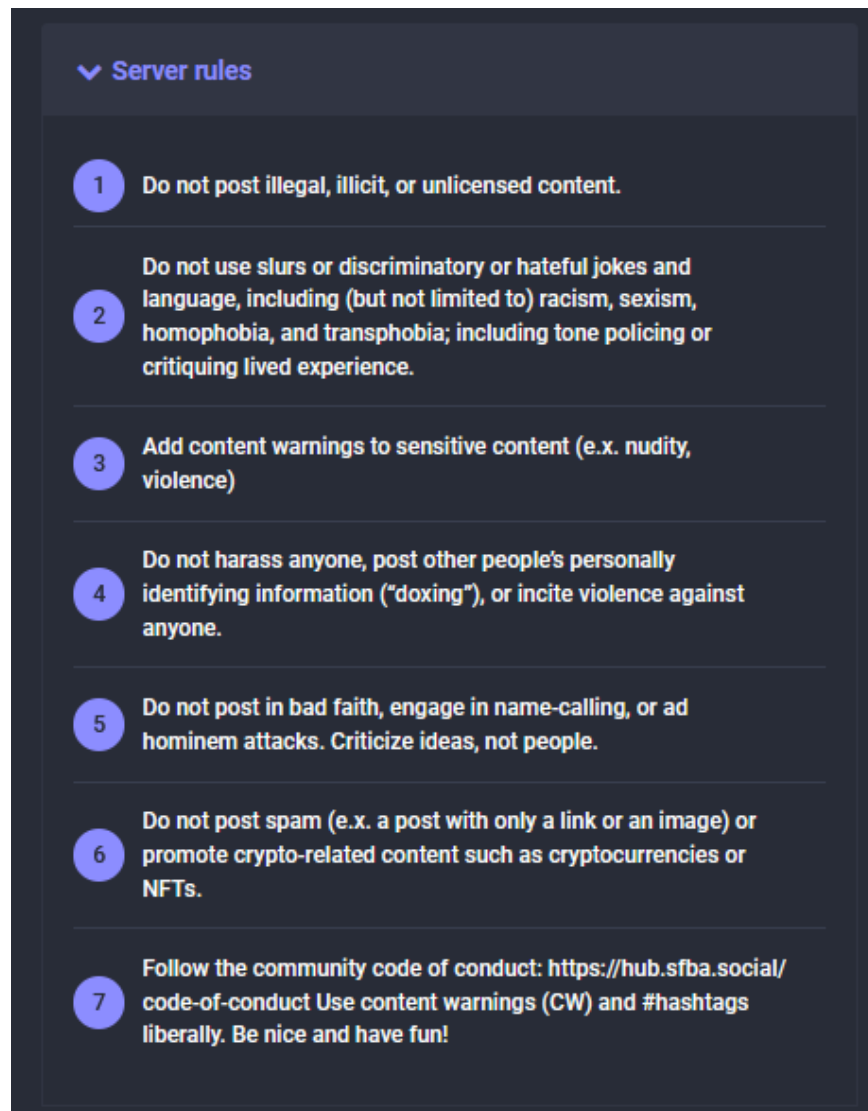


Figure 3.26: This is a screenshot of the rules on the Fediverse instance `sfba.social`. Compared to rules on Reddit and Fandom, rules here seem to be less about formatting and type of content that's allowed and more about how people behave and interact with other members.

▼ **Moderated servers**

Mastodon generally allows you to view content from and interact with users from any other server in the fediverse. These are the exceptions that have been made on this particular server.

13bells.com Reason not available	Limited
3dp.chat Spam problem	Limited
9kb.me Spam problem	Limited
aether.run Spam problem	Limited
airwaves.social Spam problem	Limited
andrew.xyz Spam problem	Limited
archaeology.social Spam problem	Limited
asbestos.cafe Reports of unaddressed hate speech	Limited
azemet.space Spam problem	Limited

Figure 3.27: This is a screenshot of the “Moderated servers” on `sfba.social`. Administrators of Fediverse instances can “defederate” from other communities, which means content can’t be shared across to `sfba.social`. The reasons for some of the moderated servers here are “Spam problem”, meaning the instance seems to be run by bots or advertisers and “Reports of unaddressed hate speech”, which means members of `sfba.social` reported the instance for hate speech but it was never addressed by the instance’s admins. Not every Fediverse instance publicly shows which other instances they’ve “moderated.”

Fediverse platforms like Mastodon are inspired by “federalist political theory” that justifies instance administrators governance actions to deny membership of users. Despite being a decentralized network of communities, Caelin (2022) showed how instance administrators came together to unanimously block the prolific hate speech community Gab that connected to the Fediverse. Work by La Cava et al. (2024) observed how polarization emerged on Mastodon from the desire to foster a federated social network that also isolates Fediverse instances that pose a risk to the stability of the Fediverse, like hate groups and spam instances. Bono et al. (2024) explored decentralized moderation on Mastodon to categorize the traits of blocked instances. Finally, Nicholson et al. (2023) introduced the first analysis of rules on Mastodon instances and how they compare to rules on Reddit. Compared to Reddit, the authors found that rules on Mastodon pay more attention to harassment and hate, reflecting the values in the “Mastodon Covenant”, and that those rules may have emerged from a lack of action on those problems on other platforms like Twitter.

3.5 Selected Empirical Case Studies and Their Rules

In this section, I briefly summarize and describe the specific subreddits, Fandom wikis, and Fediverse instances that were selected as case studies for this project, as well as their rules. I discuss the similarities and differences across these three platforms. Tables showing broad categories and examples of the rules present in the communities I studied are displayed to further contextualize the finding.

3.5.1 Subreddit Communities and Rules

We interviewed leaders from nine different subreddit communities. Community membership of subreddits in our sample ranged between roughly 40,000 to 20 million. We counted users that have subscribed to the subreddit as members of the community, although actual active participants is usually much lower. Subreddit community ages ranged between 14 to 1 year old. Subreddit communities focused on a range of topics including television and book fandoms, a support group for a chronic illness, a community for people from a specific city, a stock trading group, a general QA forum, and a buy/sell/trade space for plant enthusiasts. Leaders we spoke to either had first hand knowledge and experience of the community’s

history of rules or had become leaders of the community when the old leaders abandoned their positions and offered up their roles to existing community members. Reddit limits the number of rules that can be embedded into a subreddit’s governance system at 15, but some communities we interviewed contained additional rules on wiki pages that are linked on the subreddit’s home page. A summary of these counts can be found in Table 3.1.

Community	Created	Size	Rules
r/kdramarecommends	2014	68889	11
r/explainlikeimfive	2011	20945799	10
r/Supernatural	2008	202207	9
r/cybersecurity	2012	381366	9
r/huntsvillealabama	2008	60028	8
r/diabetes_t1	2013	41638	8
r/amcstocks	2021	67428	15
r/harrypotter	2008	1324162	14
r/takeaplantleaveaplant	2011	51156	15

Table 3.1: Reddit community characteristics. Size refers to the total number of subscribers to the subreddit.

The kinds of rules I observed in subreddit communities in our study focused on discussion behavior guidelines, content relevance and quality, legal and ethic behaviors, safety, formatting, and specific rules related to the niche of the subreddit community. I show a table of the general categories with examples from rules I pulled from subreddit’s rules’ pages in Table 3.2. Community member behaviors in discussions was almost a constant rule type across the subreddits, with rules emphasizing the importance of civility and respect of others. A main activity for subreddit members is to discuss a post, such as a news article or opinion shared by another member. Discussions are asynchronous and members are semi-anonymous. Some times, discussions may get heated as the lack of non-verbal cues and medium filter out nuance that would otherwise be present in face-to-face interactions (Walther, 1996). Discussions that deteriorate beyond heated arguments and turn into abuse are discouraged in these rules. Another major issue for subreddits is the sharing of low quality or irrelevant content. Rules are instituted to minimize low quality contributions

because it adds noise to the information that others are trying to find (Grimmelmann, 2017). Subreddits also had issues with and made rules against spam and advertising, which also contribute to the noise that make it harder to find high quality and relevant information and discussions. Finally, communities on Reddit may also institute formatting guidelines, like adding the correct “Flair” to a post, which helps categorize it, or adding spoiler tags or content warnings for content that might ruin someone’s experience with a fandom or trigger someone’s trauma, respectively.

Subreddit communities rules’ did differ in some ways. For obvious reasons, communities would have different content-specific rules related to the topic the subreddit discusses or engages in. For example, the r/cybersecurity subreddit had rules about strictly discussing cybersecurity topics for cybersecurity professionals. Some subreddits that involve trade or monetary exchanges had specific rules regulating how members trade with each other, such as r/takeaplantleaveaplant’s rules regulating member interactions when they trade plants with each other. Some communities had specific issues with misinformation, such as r/huntsvillealabama having a rule against spreading misinformation about the COVID-19 outbreak and r/diabetes_t1 having a rule against misinformation about treatment options. In summary, subreddit communities rules’ focused on regulating discussions and ensuring contributions from members are relevant and high quality, but rules diverge when communities had specific problems related to their community’s topic or specific problems that emerged in their community’s history.

3.5.2 Fandom Communities and Rules

We selected 8 Fandom communities, which ranged between roughly 900 members to roughly 23,000 members, and the ages of communities ranged between two and 16 years old. Fandoms in our sample mostly centered on comic book and anime franchises, video games, young adult fiction, and toys. For some communities, we interviewed multiple leaders who had different levels of experiences. This was the case for the DC Wiki, Marvel Wiki, Aesthetics Wiki (a fandom for fashion styles and aesthetics), and Brickipedia Wiki (a fandom for Lego hobbyists). The number of rules for each Fandom wiki community varied as some

Category	Example
Civility and Respect	“Be kind to each other, remember the human both behind and on the screen.”
Content Relevance	“All submitted content must be directly related to the TV show Supernatural.”
Quality of Content	“No low effort / poor quality posts: This is a community of cybersecurity professionals.”
No Self-Promotion/Advertising	“No advertising (includes buying, selling, or trading prescription medicine and medical devices).”
Privacy and Security	“No personally-identifiable information: Do not post personally identifiable information unless the source has consented to it.”
Moderation and Enforcement	“Moderation decisions are ultimately left up to the discretion of the ELI5 modteam.”
Spoilers and Content Warnings	“Use internal spoiler tags when discussing plot points and/or sensitive material.”
Legal and Ethical Behavior	“Market Manipulation Is Prohibited: This includes, but is not limited to, encouraging group buying, timed buying, fixed price buying, organizing, protesting, and brigading.”
Posting Guidelines	“Flair your posts correctly: For trade posts, flair your post with your country or state.”
Community-Specific Rules	“Content must be related to Type 1 Diabetes: Posts must be related to T1D at least in some way.”

Table 3.2: Subreddit Communities Rules

Community	Created	Size	Rules
Aesthetics Fandom	2018	4652	18
Marvel Fandom	2005	23481	41
Naruto Fanon Fandom	2007	4588	59
Valorant	2020	880	21
DC Wiki	2009	15025	33
Riordan Wiki	2007	13365	135
Wings of Fire Fanon	2013	10125	45
Brickpedia	2006	11926	28

Table 3.3: Fandom community characteristics. Size refers to the total number of users who have made at least 1 contribution to the wiki.

utilized additional features like Forums and instituted rules to govern those spaces while others only had rules for making contributions to their Fandom’s wiki articles. Table 3.3 summarizes the characteristics of each Fandom community in our study.

Fandom wiki community rules focused on behavioral guidelines when interacting with other community members, but also had rules related to how information is presented in the articles that members contribute to the community’s wiki. Table 3.4 shows a summary of the kinds of rules with examples of rule text from the Fandom wiki communities involved in our study. Most rules were focused on how to make edits, write articles, and format information in the wikis. Rules like this are important in peer-production communities because of the open nature of participation. Anyone can create an account and start editing a Fandom’s wiki page, but standards are often necessary to preserve the quality of the articles that the wiki produces (Butler et al., 2008).

While each Fandom wiki had general rules about civility when interacting with other members and general rules about how to make contributions, they differed in the ways that seemed tailored to needs of each community. For example, when it comes to how articles should be written, the DC wiki required an “in universe” perspective for writing the histories of its comic book characters, meaning that the content should be presented as if the events and characters are real within the fictional world, without acknowledging the real-world

creators or audience. Meanwhile, the Riordan series wiki requires that articles be written that are suitable for a younger audience because the series is marketed to young adults. How disputes are settled also differ across some Fandom wikis. Conflict disputes also varied in terms of how they're regulated, with Riordan wiki asking members to resolve conflicts on article talk pages while other wikis, like Valorant Wiki, suggested members seek out a community administrator if a conflict got too heated. In sum, the rules of Fandom wikis, while regulating interactions between members, focused more on regulating contributions to the wiki, with a lot of variation between communities as to how articles should be written.

3.5.3 Fediverse Communities and Rules

Seventeen Fediverse instance communities were selected as cases for our study. Unlike our samples from Reddit and Fandom, Fediverse instances ranged between one to five years in age. Population sizes were also different. Community sizes from our Fediverse sample ranged from as few as two members to as many as 3,700 members. Fediverse instances included in our sample may have members contribute to a particular topic or theme, but many are created for general microblogging purposes like one may do on sites like X (formerly Twitter). A summary of these Fediverse instance communities can be found in Table 3.5.

The overall rules of the Fediverse communities in our study regulate the behaviors of members in order to create inclusive and respectful spaces for their members, described in summary in Table 3.5. Generally, these communities listed rules that explicitly regulate anti-social behavior like hate speech, discrimination, and harassment so that members won't feel like they can be abused or marginalized. Rules were also instituted to regulate types of content that can be shared, with rules banning violent or pornographic imagery, spam, and advertising. Many Fediverse communities also had rules regulating against the sharing of specific ideological view points like fascism and Nazism, explicitly prohibiting any content related to them.

Rules differed in some ways between Fediverse communities. While civility and respect were norms to be enforced by most Fediverse rule sets, others had specific rules to make their communities safe for marginalized members, specifically those from LGPTQ backgrounds.

Category	Example
Behavioral Guidelines	“Be civil to other users at all times.”
Content Management	“Do not use slurs, or direct derogatory or insulting language at other users.”
Assume Good Faith	“Assume good faith; try to consider that the person on the other end is trying to positively contribute to the project.”
Avoidance of Controversial Topics	“Political discussion is not permitted on our wiki.”
Neutral Point of View	“The Neutral Point of View is a philosophy that can be best described as reporting data objectively without bias.”
Quality and Accuracy	“Articles should be written in fluent American English and from a neutral point of view in a formal tone.”
Prohibition of Plagiarism	“Plagiarism. Don’t steal content from other sites.”
Encouragement of Contribution	“BE BOLD! in updating pages. Go ahead, it’s a wiki!”
Vandalism Prevention	“Do not vandalize articles. Doing so will almost always result in a block.”
Structured Content	“We use Page Templates. The templates give a standardized look across wikis.”
Collaboration	“Be Collaborative: A wiki depends on collaboration between its users.”
Communication Protocols	“When in doubt, take it to the talk page.”
Copyright and Fair Use	“Respect copyright law. Only repost art that is free to use or with permission, and provide credit for art you use.”
Account Security	“Keep your account secure.”
Consequences for Rule Violations	“Disregard for certain rules may result in your account getting temporarily or permanently blocked.”

Table 3.4: Fandom Communities Rules

Community	Created	Size	Rules
sfba.social	2019	3700	7
toot.wales	2018	981	3
medievalist.masto.host	2018	26	4
freeradical.zone	2017	278	15
bostonmusic.online	2020	2	6
neue.city	Not Available	Not Available	22
hackers.town	2017	211	16
ausglam.space	2018	235	9
vocalounge.cafe	2020	49	6
mpsocal.net	Not Available	370	11
sonomu.club	2019	79	10
mastodon.radio	2018	1100	9
expired.mentality.rip	Not Available	Not Available	Not Available
vantil.refchat.net	Not Available	Not Available	Not Available
mk.disqordia.space	Not Available	Not Available	Not Available
glammr.us	2018	271	3

Table 3.5: Fediverse community characteristics. Size refers to the number of registered users on each instance. Note: some instances community characteristics could not be found in any public online space.

A few communities did have rules on how information is presented, like including content or trigger warnings when sharing potentially traumatic or violent content. In total, the rules of Fediverse communities appeared to regulate behaviors and content in a way that structures these spaces for open, civil discourse while also specifically calling out right wing ideologies as being prohibited.

Category	Example
Prohibited Content	“No illegal content - this server is hosted in the UK.”
Pornography and Sexual Content	“No sexual depictions of children. No holocaust denialism, Nazi symbolism, or promotion of National Socialism.”
Hate Speech and Discrimination	“Do not use slurs or racist, sexist, homophobic, transphobic, ableist or otherwise discriminatory or hateful jokes or language or promote white supremacy, anti-Semitism, transphobia, castist, or other hateful ideologies.”
Violence and Gore	“No gore or graphic violence.”
Misinformation and Pseudoscience	“It’s important that our server is a place for constructive ideas, and not ones that are used as weapons, tools of disinformation, or spreading information that can cause harm.”
Harassment and Bullying	“Do not harass anyone, participate in group harassment of anyone, or otherwise engage in personal attacks.”
Good Faith Participation	“Do not post in bad faith. That’s not nice.”
Respect for Diversity	“No racism, sexism, transphobia, or xenophobia.”
No Tone Policing	“Do not tone police.”
Mandatory CWs	“All posts with references to the following topics MUST have a useful content warning: violence; sexual content; electoral politics; self-harm and suicide; hate speech; descriptions of discrimination and traumatic experiences.”
Visual and Textual Descriptions	“All images SHOULD come with a description, using the alt-text description tools provided by Mastodon and relevant apps.”
Limited Advertising	“Advertising of shows and events are allowed. Only two posts per event please: once before the event, and once on the day of. Any more than this will be considered spam.”
Spam Prevention	“No spam, no commercial advertising, no sexually explicit or violent media, no threats of violence.”
Bot Restrictions	“No bots, unless they’re cute, funny, or useful.”
Moderation Appeals and Transparency	“Do not litigate moderation decisions on the timeline. The server provides a mechanism for appealing moderation decisions.”
Prohibited Ideologies	“If you have the following or similar kinds of ideologies or affiliations, you are not welcome here and you will be immediately banned if found out: Neo-fascism, Reactionary movements, Neo-feudalism, TERF, SWERF, or otherwise anti-trans or anti-sex worker ideologies.”

Table 3.6: Fediverse Community Rules

Chapter 4

METHODOLOGY AND ANALYTICAL STRATEGY

4.1 *Introduction*

In this chapter, I outline the methods and analytical strategies I used in collecting and analyzing the data used in this dissertation. Initially, this project was motivated by the questions of why communities often (but not always) adopt similar rules, how such rules shape the practices and experiences of online community leaders and participants, and how these processes are similar or different across several types of community platforms. Over the course of 25 months, I led, co-led, or independently conducted each phase of the research design, participant recruitment, interview, transcription, data analysis, and manuscript writing process to answer those questions, along with assistance from members of the Community Data Science Collective (CDSC).¹ Additionally, I worked to recruit and train undergraduate researchers from the University of Washington and Northwestern University on interview design and practice, as well as how to use Grounded Theory to assist with data collection and initial data analysis of interview transcripts. As the analysis of interview data progressed, the focus of analysis shifted from the initial, broad questions motivating the study to questions about the practices and beliefs that shaped out leaders adopted, enforced, and frequently changed their community rules.

4.2 *Data Collection*

Interview data collection for this project took place over several steps: creating preliminary data sets to construct lists of communities with similar rule language on each platform, identifying clusters of similar rule language communities on each platform to recruit leaders

¹The Community Data Science Collective is a multi-university research group originally based out of the University of Washington and Northwestern University. They use computational social science and qualitative methods to study online communities like wikis, subreddits, and open source software projects. <https://wiki.communitydata.science/>

from, and lastly conducting interviews. First, we assigned three members of the CDSC research group (two graduate students (including myself), one post-doc) to create data sets capturing the rules of communities on Reddit, Fandom, and Fediverse instances. I was responsible for building a data set of communities on Reddit. Reddit data came from a multiyear data collection project between 2020 and 2022 where I had been crawling Reddit’s API to record every publicly available subreddit’s rules data and storing the results in a SQL database. Subreddit communities with less than 1,000 “subscribers” (a measure that more or less corresponds to community size) were removed to ensure that some expectation of having currently active communities in the final sample, resulting in approximately 700,000 subreddits. Fandom data required analyzing archives of the edit histories of 340,282 Fandom wikis published in February 2020 on the Internet Archive.² We then limited the scope to substantive Fandom communities by including only Fandom wikis with at least 200 distinct article contributors and 10,000 article edits. Wiki data included policy pages (i.e., the rules) for community. Finally for creating a data set to sample from Fediverse instances, we used iterative sampling to discover new instances from the set of peer connections on previously known servers, starting with `mastodon.social` because of its centrality and status in the Fediverse. We then downloaded the `/texttt/about/more` HTML data on Mastodon servers and the `/static/terms-of-service.html` HTML data on Pleroma servers as they appeared on September 25, 2021. The snapshot from September 25, 2021 was chosen as this was before a new “rules” API feature was added to Mastodon, which allowed us to get the richest content for analyzing community rules. Using these HTML data, we included only Fediverse instances which had texts in English, had over 100 tokens of text on these pages (after removing stop words, articles, and non-letters).

From each setting’s data set, we conducted a clustering analysis of communities according to the language of their rules. For subreddit and Fandom wiki rules, we created a 2-dimensional encoding of each community’s rules by first concatenating all of the listed rules, reducing the dimensionality of the text by using the Doc2Vec algorithm as implemented by the `gensim` Python package, and then further reducing the dimensionality in a second stage

²https://archive.org/details/wikiteam_2020-02-09

using the U-Map algorithm as implemented in the `umap` Python package. Finally, we colored the communities according to clusters assigned using the HBDSCAN algorithm from the `hbdscan` Python package. The result gave us, in both the Reddit and Fandom setting, a list of clusters of communities based on the similarity of their rule language.

For Fediverse instances, we calculated the pairwise similarity of instances' rules by counting the number of n-grams they share in common using the following formula:

$$\frac{\text{count}(X_1 \cap X_2)}{\min(\text{count}(X_1), \text{count}(X_2))}$$

We then again used the `hbdscan` Python package to assign each Fediverse instance to a cluster based on the similarity of their rule language.

4.2.1 *Sampling Strategy*

For two weeks, we then worked to identify ideal samples from which we would recruit participants. Each empirical setting was assigned three undergraduate research assistants (RAs) and at least one graduate student researcher to manage the sampling, recruiting, and interviewing of potential online community leader participants. Because a large number of clusters and online communities existed in our data sets, we instructed the RAs in each group to make pre-liminary observations about each cluster to reduce the scope of potential online communities from which we'd recruit for interviewing based on how active and, albeit subjectively, interesting the communities in the clusters are. Specifically, RAs in each group would be assigned 10 clusters of communities in their respective empirical setting, with each third of the clusters incorporating either a large, medium, or small amount of online communities with similar rules. For each cluster, RAs were instructed to record the estimated population size of each community, whether the communities had seen activity within the previous few weeks, the kinds of rules they contained, and the general topic or theme of the community. This exercise resulted in lists of online communities we would reach out to with recruitment messages to participate to be interviewed in our study.

Before we could recruit any participants from these communities to our study, we developed protocols for both involved university's IRBs (University of Washington and Northwestern University). These protocols included study design and purpose, data collection

and storage procedures, targeted demographic range for the study, and all the supplemental material like the text of the recruitment messages, study information sheets, and interview questionnaires. More detail about these documents may be found in the Appendix. Both universities ultimately approved the study, giving us institutional permission to move forward with recruiting and interviewing research participants.

4.2.2 Recruitment

From June 2022 to August 2022, my research team recruited and interviewed a total of 40 participants from 34 different communities across subreddits, Fandom wikis, and Fediverse instances. We developed recruitment messages (see Appendix B) for each empirical setting (Reddit, Fandom, and the Fediverse) that we sent directly to moderation teams or their accounts. We used a screener survey (see Appendix D) when sending out recruitment messages to identify candidates who were familiar with rules in their communities, active participants, can speak English, and were legally adults in their country of origin. All participants we recruited were administrators or content moderators. Initial recruitment attempts were not always successful, with some communities in our data set either ignoring our not seeing or requests while others chose not to participate. While we first attempted to keep our sample contained to a single cluster, the lack of responses from some communities in the initially chosen cluster prompted us to expand our sample to other similar rule clusters in our data set. By expanding to different clusters, we compromised on being able to compare how different communities came up with similar rules, but the benefit was being able to gather more data for our study. I include quantitative characteristics about each community in each setting in Figures 3.1, 3.3, 3.5 of Chapter 3.

Participants were all offered a \$20 gift card for their participation in our study. Some participants had direct experience with creating and writing rules while others only knew the history or practice of creating and writing rules in their communities. Many leaders were around at their community’s creation and could speak to the beliefs and processes of adopting the initial rule set at that time. Other leaders were recruited onto the community’s team of moderators after the initial rules were adopted, which was the case for several leaders

on Fandom and Reddit. Some leaders on Reddit were participants of a regime change within their community, in which they were recruited to take over the previous leadership, usually because of an absence of community governance or deterioration in the quality of the community contributions (and often from a lack of any or sufficient rules). Some leaders of Fediverse instances had stated they created their community as a reaction to widespread governance issues on other, similar social network sites like Twitter. I anonymized the usernames of the leaders and refer to them in the rest of this dissertation as R, FA, and FV. For example, R3 was a leader of a subreddit community based around a popular television, FA9 was a leader of a Fandom community based on a comic book series, and FV13 was a leader of a Fediverse community based around HAM radio hobbyists.

4.2.3 Interview Methodology

I collaborated with other CDSC members on my team to develop interview protocols and questionnaires as well as a study information sheet (see Appendix A) that informed participants of the study purpose and resources they can access if they felt harmed. As the goal of this particular study was to generate new knowledge about how volunteers govern communities in different sociotechnical settings, interview questions followed a “semi-structured” format with many open-ended questions to illicit responses more likely to describe not just what participants did but also how they felt and what they were thinking and think now about situations they were in when leading their communities. The full interview protocol is available in Appendix C. The interview format started with “ice breaker” questions asking participants to reflect on their general reasons for joining and participating in their communities and what their roles were and how they generally felt about rules in online communities. Questions then narrowed down on questions specifically about the rules in their online communities, the origins of those rules and how, if ever, did they change, how rules are enforced, how their rules compare to other rules of other online communities the participant has experience with, how and why their leadership team (i.e., moderators and administrators) are organized and divide labor, and how participants felt about the rules on

the platforms that host their communities in comparison to other platform.³ The general interview protocol used for this study can be found in Appendix [# here].

From June 2022 to September 2022, we recruited and interviewed leaders of communities from Reddit, Fandom, and Fediverse communities. Interviews were conducted via computer-mediated conference call software like Zoom or `jit.si`. I trained undergraduate RAs and graduate student researchers on conducting interviews in the “intensive” style described by Charmaz (2015). Intensive interviewing relies on open-ended questions to obtain detailed responses from participants, and our interview protocol was designed to elicit detailed descriptions of participants’ experiences. As part of this training, RAs and graduate student researchers practiced interviewing on each other and were given constructive feedback before conducting interviews on participants. Per Charmaz, Interviewers were trained to follow up on ambiguous language, implicit views, and other unanticipated areas of interest that emerged in interviewee responses to help with fully exploring participants’ experiences and feelings about their work as volunteer community leaders. Each subgroup of the project was responsible for recruiting and conducting interviews in their respective empirical setting (i.e., Reddit, Fandom, and the Fediverse), and interviews were conducted by a mix of undergraduate RAs and graduate student researchers. Overall, 40 interviews were completed, their lengths ranging from 30 to 120 minutes, with most averaging around 60 minutes. In total, nine leaders from nine different Reddit communities, 14 leaders from eight different Fandom communities, and 17 leaders from 17 different Fediverse communities were interviewed.

4.3 Analytical Strategy

From October 2022 to June 2023, I led the research team on analyzing the interview data we had collected. Interviews were first transcribed with the automated transcription software `Otter.ai`. Afterward, I instructed RAs and CDSC members to go back over each interview to correct transcription errors that occur from faulty microphones, quiet speakers, or too much feedback. Corrected interview transcripts were then assigned to three groups within

³The last set of questions was to help understand how the sociotechnical design influenced, if at all, their experiences of rules in online communities

the research team, each led by one graduate student, including myself, and two to three undergraduate students, to conduct line-by-line or event-by-event initial coding with the open source software for qualitative coding Taguette. This initial coding process requires examining the data by going through each line of the interview transcript and generating a “code” that captures the actions and feelings of what the participant is saying. This process helps break apart the text to allow researchers to discover emergent themes and helps to avoid imposing preconceived notions on the data. I trained undergraduate research assistants on performing initial coding from the strategies for constructing Grounded Theory outlined by Charmaz (2015). I met with undergraduate research assistants each week for several weeks to evaluate initial coding performance and discuss preliminary emergent things in the interview texts. Approximately 4,500 initial codes were created through this process. Following through on the methods outlined in *Constructing Grounded Theory* (Charmaz, 2015), we needed to be able to group the initial codes in some way to be able to efficiently compare the context in which they’re deployed and discover patterns in the interview data. To do this, I had each research assistant label their initial codes with a broader category. Hundreds of categories emerged out of this process, but many overlapped and needed to be consolidated into other categories. I worked with two other CDSC researchers to consolidate and merge the number of categories down to 20. Transcription corrections and initial coding passes took approximately 4 months, ending in January 2023.

From January to February of 2023, I started the next step of my Grounded Theory analysis by generating “focused codes.” Focused codes involves choosing initial codes “which make most analytic sense to categorize the data” and creating labels for them that move up to a conceptual level (whereas initial codes are more grounded in the data) that can help be used for synthesizing a large amount of interview data (Charmaz, 2015). I assigned to myself and two CDSC researchers a set of categories of initial codes for generating this next level of focused codes. To focus our analysis on the research questions guiding this study, focused codes were created for 11 out of the 20 categories: *rule adherence and enforcement, governance strategies, governance challenges, governance communication and feedback, governance values and philosophy, rule creation, rule change, perceptions and experiences of community rules, technological challenges and affordances, rule communication, and moder-*

ation team recruitment. This exercise resulted in 337 focused codes. From March to April, 2023, in collaboration with another CDSC researcher, I used Miro first to visualize relationships between focused codes and the broader categories with a focus on rule creation, rule change, governance values and philosophy, governance strategies, technological challenges and affordances, and perceptions and experiences of community rules (see Appendix ??). After organizing focused codes and grouping them into categories, I analyzed themes between categories and conceptualized a process of governance, shown in Appendix F. By May 2023, we both drafted several research memos related to rule creation, rule change, and sociotechnical differences for each platform. I met regularly with other CDSC researchers to discuss and iterate on the ideas, themes, and concepts from the memos.

One initial challenge was finding sufficient evidence across interviews because, although we had some oversight over undergraduate research assistants, initial codes were not always applied consistently or as often as possible in interview transcripts. To solve this issue, I created a code book using 42 themes from the initial research memos as codes and applied them as often as possible to all 40 interviews. We treated this process as a version of “thematic sampling” in which the researcher seeks out evidence in support of thematic codes and concepts from other interviews (Charmaz, 2015). I organized these into multiple broad categories: community, initial rules, leader communication, problems, rule change, rule communication, rule feedback, rule purpose, rule writing, and sociotechnical differences. Interview quotes were re-evaluated this way, and new memos were produced that iterated on themes from the initial memos. During this process, I moved from focused codes to interview data to theory and back to generate empirically grounded contributions to the theoretical problems that motivated this study, such as why groups adopt similar rules, how groups successfully enforce rules, and when and why do groups eventually change their rules. The research team met repeatedly to give feedback on memos and new iterations of these ideas until the findings could be distilled into the concepts in this manuscript.

Chapter 5

RULE ADOPTION: THE FUNCTIONS OF RULES IN ONLINE COMMUNITY GOVERNANCE

5.1 Introduction

The initial questions motivating this study focused on why and how online communities adopt rules, why similar rules emerge, and how these processes differ across platforms. However, interviews with community leaders revealed that rule adoption is a more complex, multilayered process than originally believed. While rules often address immediate behavioral concerns, such as defining acceptable interactions and content, they also serve broader functions: *reinforcing community identity* and *signaling legitimacy to external stakeholders*. In the following sections, I outline these three different functions of rules (see 5.1 for a brief summary) by discussing the specific problems they address, the conditions under which leaders implemented them, and how these functions influence communities' overall success.

Rule Function	Community Problems	Implementation Process	Community Outcomes
<i>setting behavioral expectations</i>	Openness Problematic behaviors Community growth	Find compatible rules from existing communities Problem-specific rule addition / change	Foster a safe space Moderate member-member behaviors Prevent interpersonal conflicts
<i>reinforcing community identity</i>	Openness Ambiguous goals Newcomer noise	Precise definitions in rules Specifying the intended audience Redirect newcomers to appropriate space	Strengthen identity boundaries Sustain community ecosystems Retain high quality members
<i>signaling legitimacy to external stakeholders</i>	Liability of newness Community interdependence	Copy rules from high status communities Pledge to cooperate with standardized rule set	Retain access to external resources Overcome newness: survival Reinforce broader social norms

Table 5.1: This table summarizes the problems, implementation, and outcomes of the three different ways rules function in online community governance.

5.2 Setting behavioral expectations

Rules are initially adopted to solve well known problems inherent to managing online communities, namely member incivility and low quality or harmful contributions (Grimmel-

mann, 2017). Leaders of online communities we spoke to want their communities to be spaces of safety and inclusion for their members, and rules reinforce leader authority to moderate the behaviors of the community’s participants. This is what I refer to as a primary function of rules in online communities: *setting behavioral expectations*. In this section, I describe the ways in which leaders thought about adopting rules that solved the initial problems of managing online communities and the different strategies they took in adopting their initial rule sets.

5.2.1 Community Problems

Leaders of online communities primarily moderate the behaviors of the users that participate in order to “facilitate cooperation and prevent abuse.” (Grimmelmann, 2017) The openness of online communities invites users of both different degrees of knowledge for a given topic and cultural expectations about how to interact with other people online. In other words, while someone with expert knowledge may have valuable contributions, their negative interactions with other members could contribute to a toxic environment. One participant, FA4, described the distinction between managing information content shared by their community members and managing their interpersonal behaviors as having to “find the balance between...fostering a healthy community, ensuring we have the best editors” without prioritizing “the capacity of the editors over their behavior.” Interpersonal conflicts in online communities may cause other community members to participate less or leave the community entirely. FV7 and FV11, for example, both left Twitter and created communities in the Fediverse because Twitter was “too toxic” (FV7) and because “there was a lot of abuse and harassment” (FV8) going unaddressed by Twitter administrators. One of the biggest problems for communities was how to handle antisocial behavior, harassment, and abuse.

Leaders across all three platforms claimed as their reasons for having rules was to make their community a “safe space”, promote civility in member interactions, and prevent harm. For some, this was their primary reason for having rules at all. For example, R9 manages a subreddit community for people to trade and care for house plants. When asked about the

purpose of their rules, they said:

“I think the biggest purpose of the rules is to keep everybody safe. Make sure nobody’s getting scammed and then the like secondary rules of like, be kind and no harassing and...make it so that the community doesn’t become a toxic place as well.”

The safety here first refers to preventing members from being “scammed” by malicious actors, but the second reference is to interpersonal behaviors, ensuring members are civil when communicating with each other. FV10 said in regards to why their community has rules, “I think that it’s pretty important in terms of setting the vibe...sort of creating expectations for people who might want to join.” If content moderation policies advertise the experiences one has on platforms (Gillespie, 2018), then, similarly, rules in autonomous online communities are a guide behavioral expectations from community members. Having identified the challenges leaders faced in managing member behavior, the next step is in understanding how they implemented rules to address those problems.

5.2.2 Rule Implementation

Rules were adopted in different ways. Although a couple Fediverse participants said they created their initial rules from scratch, many other interview participants stated their initial community rule sets were either directly imported from or influenced by another community’s rule page. FA14 stated that community leaders are “always inspired by other communities” and that they “always take a template of rules” from other communities and “basically repeat rules like **Don’t use swearing** or **Be nice to each other.**” FV10 rationalized the process for selecting their initial rules by relating it to how they legal documents: “you look at other examples of how other people have done it, and then you copy the good stuff and you sort of build on it.” What counts as the “good stuff” was subjective. According to FV12, finding the right rule set in other communities meant those rules would be “the things that I would say...these are the things that I would fight for.” For these participants, selecting the initial rule set meant finding a set of norms that align with their own beliefs about how online community members behave and participate.

That said, not all rule sets were simply copied and pasted into a rules page, and many had to be tailored to a community's more specific needs. Participants like FA4 described this process with "we didn't just copy the rules...we saw what exactly worked for us, what didn't, and what we could ask that the other sites were not asking, and we adjusted to our specific community." Other participants assembled their rule sets from a variety of other communities' rules, picking and choosing rules to solve a specific problem regardless of their overlap in topic. For example, FA11 claimed that their policies on editing in their wiki (a Fandom for a book series) came from another Fandom for Avatar the Last Airbender (a Fandom for a television show). Another example comes from R4, who needed a rule on how people should advertise in their community. They "basically just copy and paste[d]" a similar policy on excessive promotion of one's product or social media from the "Reddiquette", which is an informal set of guidelines created and shared by subreddit leaders.

Rules were also adopted over time. Several Fandom participants claimed their community had no rules at its onset. One reason was because in the beginning, their community was small and not that active. FA5 said that rules didn't seem necessary because their community began as just a wiki for a small group of friends. FA6 also rationalized the lack of an initial rule set for their wiki by saying:

"the number of edits was in the very low volume...so the number of times that people would break rules, either intentionally or mistakenly, it was really low. So we just dealt with it as a team, but then we ended up adding [rules]."

In these cases, leaders believed that enforcing community norms could be achieved without formalizing these norms into written, enforceable rules. However, unaddressed behavioral problems may threaten the success of an online community when they persist, leading to the deterioration of what the community provides, such as an accurate information repository on a Fandom wiki or a safe space to discuss a topic on a subreddit or Fediverse instance.

Growth was one factor that could exacerbate this problem. R9, despite already having a rule set, had to add rules to deal with an influx of newcomers, some of which were trying to harm community members through scams:

“Since the community was growing so quickly, there was an influx of scams and stuff that were happening, and so there was lots of new rules being implemented.”

In this example, R9’s leadership team implemented a new rule requiring members go through a verification process to participate in the activity of selling and trading plants. Before, R9 felt that “almost all [subreddits] force people to use a goods and services [verification rule]” but that their community had held off from copying this rule from other communities because “it is more annoying for sellers.” For R9, this was an ongoing problem, but it didn’t become nearly as bad as when their community grew, triggering the need to implement a solution with a rule. Growth also affected wiki communities. FA3 said their community was seeing “so many troll accounts, so many vandal accounts, so many people sock-puppeting that it just became too much of a hassle to deal with” and subsequently implemented a “four day edit rule” that prevented new accounts from interacting with the community until four days after account creation.

Conditions for Rule Implementation

Drawing from my interviews, I propose some conditions where formal written rules of an online community may emerge. First, leaders who create a space for a community may arrive with previously gathered experiences and expectations about how people should interact with each other in an online community. These expectations may be formalized as written rules at the time of community creation. Leaders acknowledge the potential for incivility or disruptive behaviors and proactively inscribe formal authority to sanction those behaviors before any have ever happened. This action sets behavioral expectations from the start, hoping that newcomers read and embody the rules before participating. These rules are usually implemented first through a search for existing rule sets from other online communities. Some rule sets are adopted word for word into one’s community while others are pieced together from various other rule sets or modified to better fit one’s community goals or topic.

Some leaders may add rules after the time of community creation. Rules may emerge as the resources to informally sanction behaviors doesn’t scale with growth in membership.

More participants invites more opportunities for misunderstandings and a wider range of topic expertise and motives. Some behavioral problems may even have been tolerated for a time, but an influx of these problems can cause the community to reach a breaking point and a need for formal sanctioning. Here, rules are implemented (or changed) to address a specific problem that can no longer be tolerated in the community.

5.2.3 Community Outcomes

The desired outcomes for setting behavioral expectations with rules is fostering a safe space for members to participate which helps for attracting and retaining prosocial, productive members, contributing to community success. Rules help regulate the interactions that occur within a community space and achieve community goals, as I show in Figure 5.1. However, this also results in more work by leaders to enforce rules with sanctions. Sanctioning rule breakers adds an additional cost to imposing a negative externality on a community, and such costs may drive potential disruptive or antisocial actors away from harming community members. Moving beyond the immediate purpose of rules within individual communities, the next two sections delve into abstract, secondary functions of rules that regulate inter-community dynamics. Specifically, I explore how communities dependent on each other may adopt similar rules to reinforce broader norms and signal legitimacy, while those in similar niches may differentiate themselves through unique rule sets to assert and maintain distinct community identities.

5.3 Reinforcing community identity

While rules may initially be adopted for preventing and solving problems related to managing incivility and low quality contributions, leaders often used rules to define their community's core identity and differentiate it from other communities. Openness of online communities, while beneficial for community growth (Kraut et al., 2012), can become a problem when members don't understand the community's goals or arrive with misaligned expectations about how to engage with other community members. In these instances, rules serve to clarify the community's intended purpose, with enforcement helping to police the boundary between what the community is and what the community isn't.

5.3.1 *Community Problems*

Interviews showed that rules not only shape interactions within communities but also regulate communities' relationships with others in a broader social network by reinforcing a community's identity. Online communities benefit from their openness—the ability for newcomers to freely participate helps the community grow and become more productive (Grimmelmann, 2017). That said, newcomers may not clearly understand community goals, and their participation, civil or not, may disrupt communities with sensitive group norms. This was especially prominent in communities that occupied a particular niche topic. Online community niches are sustained by the existence of smaller communities that specialize in an aspect of the topic and attract experts to help share useful information (Hwang and Foote, 2021; TeBlunthuis et al., 2022). Off topic questions or incendiary topics can threaten the stability of an online community by diluting the provision of focused expert information and undermining the purposeful engagement that defines the community's core identity. When this problem persists, leaders may adopt or refine rules that reinforce the boundaries of the community's identity, ensuring its purpose and group cohesion remain intact.

Interviews revealed that rules shape not only interactions within communities but also regulate how communities relate to others in a broader community network by reinforcing their identity. While openness allows newcomers to participate and support community growth and productivity (Kraut et al., 2012; Grimmelmann, 2017), it can also lead to misaligned expectations. New members may not fully understand the community's goals, and even well-intentioned participation can disrupt sensitive community norms. This was especially true in niche interest communities, where the focus on specialized topics relies on attracting experts who contribute valuable information (Hwang and Foote, 2021; TeBlunthuis et al., 2022). Off-topic or provocative discussions may dilute the contributions of expert knowledge and undermine the community's goals and purpose. To protect the community's core identity and maintain group cohesion, leaders often adopted or refined rules that reinforce group boundaries and ensure the community's goals remain clear and intact.

5.3.2 Rule Implementation

I saw this more abstract function of rules exemplified in spaces like R4's community, which is designed to be a space for cybersecurity professionals working in the industry to discuss the latest news and technology related to their work. R4 described issues with newcomers frequently asking for personalized cybersecurity help, as if the `r/cybersecurity` subreddit is a Q&A technical forum. Despite being called `r/cybersecurity`, R4 did not intend for the community to be a place for personal troubleshooting questions but rather a space specifically for cybersecurity professionals. One rule was implemented to define the identity of the subreddit community:

Rule 2: Must be relevant to cybersecurity professionals

Questions about personal security **MUST** be posted on `r/cybersecurity_help`.

This includes:

- How to stay safe online
- If you think/know you've been hacked
- How to secure sensitive info
- Analyzing suspicious links or emails you received
- Etc.

Similarly, use `r/homesecurity` for physical security questions (ex. burglar alarms, security cameras).

This does not mean “no questions.” Cybersecurity questions by technical professionals & students for technical professionals are strongly encouraged.

R4's rule reflects an attempt to define *who* this community is for and identifies other community spaces on Reddit to participate in when one's contributions or experiences don't fit the criteria for `r/cybersecurity`. R4 justification for this rule was, “we're trying to create a community for technical professionals, a lot of that personal help doesn't really work out, so we created the `[cybersecurity_help]` to split up those questions.” Another

leader we spoke to, R2, explained how the rules of the two similar communities they help lead were implemented to draw strict boundaries between them and intentionally structure different means of participation:

‘`r/explainlikeimfive` has a very strict and rigorous environment — as rigorous as you can sort of be within the confines of Reddit. It has fairly strict rules on how you answer questions and what types of questions can be asked. They’d have to be objective, complex, nothing that asked for opinions or speculation or anything like that. `r/nostupidquestions` [is] significantly more freeform. The idea is to allow as many good faith questions on pretty much any topic.’

From this explanation, R2 explains that their rules structure different experiences in two communities that overlap in the theme of general Q&A. This is important because the existence of both communities is necessary for either one to be successful. One community, `r/explainlikeimfive`, preferred to specialize in questions that would have “objective explanations” and not “any question you may have.” Additionally, responses to questions in this community must be “friendly, simplified and layperson-accessible explanations.” The other community, `r/nostupidquestions`, preferred to generalize to allow any topic and question to be asked because “everyone has questions that they may feel uncomfortable asking certain people, so this place gives you a nice area to not be judged asking about it.” The relationship between these two communities becomes clear—with the existence of `r/nostupidquestions`, “any question you may have” can find its way to an audience who may be willing to help answer, therefore helping filter out non-objective questions and unhelpful responses from the `r/explainlikeimfive`. The “strict and rigorous” environment that R2 described is made possible by the enforcement of the rules that define its identity as a more serious, objective space for learning.

One example of this also emerged in an interview with a forum moderator of the DC Wiki. This participant, FA9, stated that the goals for the forum part of their Wiki community was to have “meaningful discussions” but that this wasn’t always clear to people who visit, some of which wish to post off topic threads or fan fiction related to the DC comic universe. FA9 described a problem in Fandom wiki more generally about how fan fiction is organized

in relation to a Fandom’s canon story lore—that this is “confusing...even when I like fan fiction...I don’t believe that’s the place [the forum] to put it out.” FA9 drew a distinction between the DC Wiki’s forum rules and the Marvel Wiki’s more “permissive” forum rules, which actually allow content like fan fiction and off topic discussions. Despite both being wiki communities for two of the most popular comic book franchises in the world, the DC Wiki leaders feel that on their community, participation should be objective and relevant. FA9 describes this goal, saying:

“It’s not just to probably have people participating for just the sake of participation, you can do that somewhere else. It’s just that [DC Wiki admins] want the discussions engaging the comic book stories more than media itself.”

FA9 brought up the Marvel Wiki’s rules throughout the interview to use as a comparison against the DC Wiki’s rules with respect to how much more strict and “limiting” the DC Wiki rules for forum participation are than the rules on the Marvel Wiki. In this example, FA9 is describing the experience of an individual’s relationship to the online community. Visitors may have misaligned expectations about acceptable topics, which can clash with the community’s established goals. Restricting content related to fan fiction reinforces the DC Wiki’s forum as a place for not just serious, meaningful discussion but also unambiguously relevant content in the Fandom’s lore.

Fewer participants from the Fediverse communities explicitly discussed how rules were used to shape their community’s identity, although on a few occasions it was implied by comparing one’s own Fediverse instance to those they labeled “free speech” communities. FV13 described these communities as places where its leaders “don’t care about moderation—say what you like” and actively blocked those servers from being able to connect to FV13’s Fediverse instance. Such “free speech” servers implied a community where moderation on behavior and content is virtually unrestricted but at the cost of abuse to community members. FV8 stated their rules are in place to “signpost” to outsiders that this is not a space for a “free speech warrior” because FV8 is “interested in everybody having a good time and being nice to each other, and if that infringes on your ‘free speech’, well, sorry, you can go somewhere else.” Indeed, FV14 echoed a similar belief that outsiders may see their

community is not a “free speech” and self-select out of registering to join FV14’s community:

We’re a small community. People don’t usually join this instance, and they just see on the about page, it’s not a ‘free speech’ instance. This is not...we do not tolerate any kind of discrimination...They probably do not join after it. Well, there’s better instances for this on Fediverse.

FV14 being both a small community and stating there are “better instances for this” implies that 1) their leadership is not interested in growth if it’s at the cost of disrupting the shared group identity around norms of civility and 2) the existence of Fediverse instances that do align with a “free speech” user means that ultimately that user can find a suitable community.

Conditions for Rule Implementation

I propose several conditions under which a community’s identity may be jeopardized, emphasizing the need for rules to clarify and reinforce goals and a shared identity. For instance, an influx of external actors (i.e., newcomers), may misunderstand a community’s purpose, especially when its goals appear ambiguous. In the case of `r/cybersecurity`, the community’s name may suggest it is open to a general audience interested in cybersecurity. However, its leaders intend for it to cater specifically to cybersecurity professionals. Contributions from non-professionals can dilute the quality of information and discourage experts from participating, ultimately threatening the value that the community provides. As non-professionals begin to outnumber the core group of experts, the community’s identity risks being drowned out by newcomer’s off-topic noise. To address this, the leaders of `r/cybersecurity` implemented rules that clearly define the target audience, specify who the community is not for, and outline acceptable types of discussions. These rules served to protect the community’s group identity and maintain its focus on its niche interest. In other words, the need for some rules may arise when communities seek to specialize, rely on a cohesive group identity for member retention, and balance openness to newcomers reinforcing and maintaining the community’s intended purpose.

5.3.3 *Community Outcomes*

By implementing and enforcing rules in this way, community leaders created an environment that encourages productive and relevant contributions. Rules that reinforce community identity allow communities that specialize in a niche topic to distinguish themselves from more general interest communities that may overlap in the same topic area. This differentiation is crucial for sustaining the ecosystems of online communities, as it helps ensure that specific or general audiences can find the appropriate space for their needs (Hwang and Foote, 2021; TeBlunthuis et al., 2022). Additionally, clearly articulated rules provide newcomers with a clearer understanding of the community’s goals, helping them to self-select into spaces that align with their own expectations.

5.4 *Signaling legitimacy to external stakeholders*

In trying to understand why communities adopt similar rules, I found one explanation in the observation of leaders copying rules of more successful communities to signal that their community is legitimately governed. Leaders of new communities often challenges in gaining legitimacy, especially within established networks of other online communities. Their community may be perceived as less credible due to its newness, and inexperienced leaders may feel insecure about their ability to write formal rules or effectively govern. Implementing rules modeled after trusted, established communities served as a way to signal legitimacy to external stakeholders, such as potential newcomers or leaders from other communities.

5.4.1 *Community Problems*

New communities often face the same “liability of newness” that threatens the survival of emerging organizations. This concept, rooted in organizational sociology, suggests that younger organizations are more likely to fail than older ones, largely due to their lower levels of legitimacy, which restricts their access to crucial resources in their respective fields (Singh et al., 1986). Similarly, new online communities often faced struggles with inexperienced leadership and a lack of governance or rule-making expertise. These communities rely

on attracting a broader audience from their platforms to grow their membership, which can be difficult without some established credibility. In the Fediverse in particular, new communities were often particularly dependent on positive relationships with established other communities to gain access to their audiences and content. My analysis here shows that leaders may often adopt similar rules from trusted or high status communities to signal legitimacy and secure the external support they need for survival.

5.4.2 Rule Implementation

Many participants described in their process of rule creation a search for existing rule sets to either adopt as their own or adapt to their community’s specific needs. This process was often guided by a feeling like they were not “actually experienced enough and knowledgeable enough” (FV12) to create rule-sets themselves. However, during the rule creation process, leaders not only sought rules that align with their desired behavioral expectations within the community but also considered how the rules they implement would shape external stakeholders’ perceptions of their community and ability to govern. In other words, as FV10 explained, “that’s an important function of the rules...to send signals to other admins and that they know what we are moderating.” In interviews, I identified two approaches leaders used to signal legitimacy to external stakeholders: copying rules from more established communities or committing to standards of community guidelines widely shared by a professional network of community leaders.

For example, some Fediverse leaders we spoke to initially adopted the same rule-set from the ‘`Mastodon.social`’ instance. This instance was created by Eugen Rochko, the developer of the widely used Mastodon social networking platform that so many Fediverse instance admins rely on for connecting to other communities. FV5 described the logic behind selecting the rules of this Fediverse instance for their newly created community:

“So I needed I needed a baseline for rules. And so I just went to the largest instance, and I think Mastodon.social was hosted by Eugen. It’s hosted by Eugen right, like, he’s the guy who, you know, if you’re gonna, if you’re gonna copy rules about a new community, you might as well go right from the source.

So I copied most or all of the rules.”

Here, FV5 trusts that adopting the rules from **Mastodon.social** would benefit their own Fediverse instance, citing **Mastodon.social**’s large community size and high status of its creator. Another participant, FV2, acknowledged the prominence of **Mastodon.social** within the Fediverse, despite having limited knowledge about its creator. They detailed their process for adopting its rule-set, stating:

“I went to other instances, I looked at what they had in terms of rules. And the rules from the **Mastodon.social** instance made sense to me. And I also saw a lot of other instances referring to those rules. Which kind of like created this network effect.”

As a leader of a new community in the Fediverse, FV2’s concluded that not only does **Mastodon.social** have a high legitimacy within the Fediverse because of its size and status but also because of its high degree of centrality in the network of communities on the Fediverse that reference **Mastodon.social**’s rules. In both cases, the size and status of **Mastodon.social** led participants to view its rule-set as highly legitimate. FV2 believes that external users “likely look at the server rules, and review them in terms of like will those rules protect me sufficiently?” By adopting the **Mastodon.social** rule-set as their own, participants hoped to make their own communities seem more legitimate to others. This even occurred with a larger, already established community, the **Marvel Wiki**, where leaders like FA3 “compare [their community] to the other big wikis and see how they address these kinds of issues and go from there.” In all of these examples, community leaders draw rules from other, highly regarded communities, seeking to emulate their governance practices to enhance their own legitimacy.

Another approach for establishing high legitimacy was observed in participants who pledged to cooperate with a standardized set of rules shared across a network of online community leaders. One example was the pledge to the Fediverse-Friendly Moderation Covenant (FFMC), which FV6 characterized as “sort of a standardized set of rules to ensure greater cooperation in the Fediverse.” The FFMC is an unofficial document shared on a

GitHub page that is primarily meant for instance admins on the Fediverse. It’s maintainers hold that the FFMC isn’t meant to “replace an instance’s set of rules” but rather should be an extension to their community’s rules that others should agree to follow so that “the networks integrity and the ability of users to interact with as many people as possible is kept intact.”¹ The core rules of the FFMC document reflect an attempt to hold the decentralized communities on the Fediverse together:

1. Administration and Moderation will be handled as neutral as possible.
2. Actions taken will respond to the bare minimum to keep the network operating and not influence the ability of the instance’s users to connect to other people.
3. Communication before Actions.
4. We won’t recommend blocks.
5. No Guilt-by-Association/Federation.

In fact, most Fediverse administrators we interviewed discussed governance in terms of moderating content and users that appear in their community’s timeline from *external* Fediverse communities. Sanctioning occurred mostly on other communities, such as muting the external instance from appearing on one’s own instance timeline or blocking their ability to view and interact entirely. However, blocking too many other Fediverse instances undermines the success of a defederated social network. If every community has blocked each other over perceived problems, then they all suffer as they’re cut off from more audiences and information. With these rules, the FFMC attempts to minimize the amount to which Fediverse communities will decide to block each other, calling this action a “last resort.” That said, some participants, like FV1, justified their adoption of the FFMC because they wanted to be seen by other instance administrators as “a stand up citizen—a stand up member of the Mastodon community” and also because it put them in a “bubble of known instances.” FV1’s adoption of the FFMC not only signals its commitment to a successful

¹<https://github.com/bildpunkt/fediverse-friendly-moderation-covenant>

federated decentralized social network to the wider Fediverse, but it also is perceived to gain legitimacy by aligning with other high status communities that also endorse the FFMC.

Another example came from not copying specifically rules but adopting “blocklists” of known instances that usually don’t moderate content and prevent abuse from already established communities. FV5 recalled their reasoning for adopting the blocklists from **Mastodon.social**:

“I copied the blocklist from I think **Mastodon.social**...I didn’t want to be federated with Nazi adjacent Mastodon instances. So I copied a sort of generic block list.”

FV5 makes the assumption that **Mastodon.social**’s blocklist is legitimate because of its status in the greater Fediverse community network and made the decision to block the same Fediverse instances. Others like FV12 refer to a list of harmful communities that appear on a continuously updated “fediblock” article that lists Fediverse instances that generally allow harassment and hate speech toward marginalized groups. Over all these examples, new leaders put their trust in already established community leaders decisions and make similar governance decisions to increase their community’s own legitimacy.

Conditions for Rule Implementation

I identify some conditions that explain the widespread similarity of rules across online communities and why leaders adopt them in this way. First, inexperienced leaders, uncertain about how to write effective rules or unaware of what is socially acceptable within their environment, may often copied rules copy rules from communities they perceive as successful. Second, new communities that rely on external resources—such as access to a broader audience or new information or content—may adopt rules from established, legitimate communities in an attempt to enhance their own credibility and improve their chances of survival.

5.4.3 *Community Outcomes*

By adopting rules this way, new community leaders may achieve greater perceived legitimacy, gaining the social capital needed to access external stakeholders, as members and content or information from other communities, as well as to attract newcomers and grow their own membership. This approach may help communities overcome the liability of newness that can often lead to early failure. However, simply copying rules—particularly in Fediverse communities—did not guarantee that those rules are fully embraced and enforced internally by leadership. They may be symbolically adopted to appear legitimate but mere adoption doesn’t guarantee enforcement.

5.5 *Discussion*

In this section, I present the rule function conceptual framework to understand the ways that online governance is enacted, from online communities to broader digital ecosystems and platforms. I argue that a primary function of online community rules is to regulate behavior and moderate content. The establishment of formal written rules gives leaders legitimate justification to enforce the norms that help their communities achieve their goals. I’ve also shown how rules can take on two additional functions beyond their initial formulation to help communities thrive: reinforcing identity and signaling legitimacy to external stakeholders. I illustrate this conceptual framework of rules function in Figure 5.1 and discuss the implications of these other two functions in the sections that follow.

5.5.1 *Managing Identity Boundaries through Community Rules*

My findings reinforce an idea from previous work that community leaders may intentionally define their community through the adoption of rules (Fiesler et al., 2018). Decisions to reinforce a community identity are often influenced accompanied by the acknowledgment of problems in the community that aren’t just disruptive to day-to-day functioning but also threaten the success of the community’s goals. Community leaders try to make sense of their organizational environment and what these problems mean and how and when to act (Weick et al., 2005). These interpretations contribute to the construction of the group’s

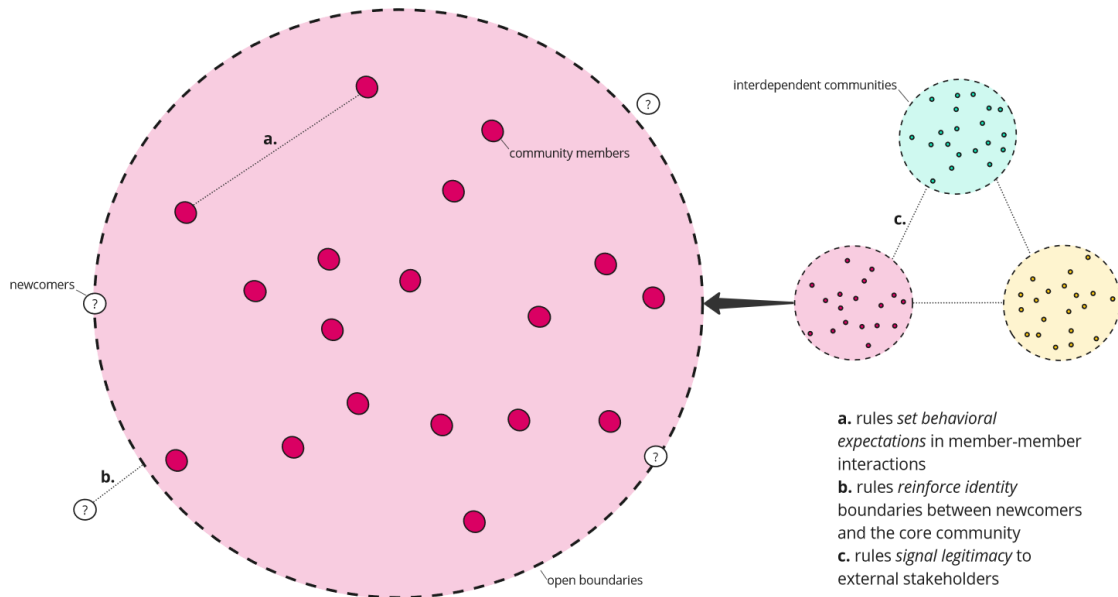


Figure 5.1: Illustrates the three different functions rule may take in the management of online communities.

organizational identity, which “helps members make sense of their situation by clarifying the defining attributes and purpose of the organization, thereby reducing ambiguity and providing direction.” (Santos and Eisenhardt, 2005, p. 500) For online communities, this identity construction occurs in the establishment of formal, written rules.

Online communities face a unique environment. Take Reddit, where new communities emerge from existing other communities (Tan, 2018), and membership often overlaps (TeBlunthuis and Hill, 2022; TeBlunthuis et al., 2022) as users wander around communities (Tan and Lee, 2015) given the openness afforded in Reddit’s platform to participate in virtually any public community space. When attention is drawn to a community and it encounters a massive influx of newcomers, leaders may evaluate and strengthen their rules to reinforce norms and the community’s identity (Kiene et al., 2016). Leaders of online communities use rules and their enforcement as a means of policing the boundaries between their community space and everything outside of it as well as socializing newcomers into the community’s shared identity.

These findings show that decisions to reinforce group identity through rule implementation are often made through the analysis of how one’s community fits within a broader ecosystem of similar online communities. Prior work by Hwang and Foote (2021) and TeBlunthuis et al. (2022) has shown that seemingly similar online communities often provide different purposes and complement each other rather than compete. This ecology approach to online communities asks scholars to shift “the focus from environmental forces to commensal relationships between communities that can shape each other’s growth.” TeBlunthuis and Hill (2022) Here, rules play an important part in sustaining what Hwang and Foote (2021) call “nested niches”, or small communities that specialize in a topic or goal but rely on their relations to larger communities in an “embedded, complex, symbiotic socio-informational ecosystem.” Communities may benefit from remaining small, as it affords a more controlled space that can better restrict the kinds of behaviors and content allowed while staying connected to larger communities that help sustain membership churn (Hwang and Foote, 2021). This kind of control allows a community to specialize in its goals (usually by narrowing on a topic), thus providing an important source of information in this larger, “socio-informational ecosystem”. People who identify with a specialized topic are likely to participate if they sense that their identity aligns with the group’s (Chatman, 1991). Therefore, the enforcement of rules not only strengthens a community’s distinct identity but also delineates the boundaries between different community spaces, facilitating a more efficient distribution of informational resources and community engagement within a socio-informational ecosystem.

5.5.2 Legitimacy and Cooperation in the Governance of Online Community Networks

As rules reinforce a community’s identity internally, they also project legitimacy to external stakeholders, helping communities secure resources and build trust in a larger community network. One may find it puzzling that leaders in the decentralized community network of the Fediverse felt compelled to cooperate with broader, collective norms by adopting similar rules and policies. We might expect each individual community in the Fediverse to establish its own “fiefdom” that rules autonomously, as they are not technically tied to any central

authority. The design of sociotechnical systems hosting communities that facilitate such arrangements has become of increasing concern to social computing scholars. Schneider (2022) has argued that the proliferation of “implicit feudalism” permeating the governance of online communities is a result of sociotechnical designs encouraging hierarchy and authoritarian social organization (Schneider, 2022). With specific regard for communities on the Fediverse, Nicholson et al. (2023) encouraged future research to investigate the arrangement where no central authority has power over any one community and how this kind of community autonomy might lead to “benevolent dictators for life” that encourage similar “toxic technocultures” that permeate communities on Reddit (Massanari, 2017). This institutional arrangement is worth reflection. Nicholson et al. (2023) are right that instances on Fediverse platforms like Mastodon “are not accountable to a higher platform authority in the same way that subreddits are,” but these findings reveal a horizontal force toward cooperation. I propose that normative isomorphic pressures may emerge in a decentralized community ecosystem that can lead to many communities either copying rule sets of more prominent status communities, pledging to adopt and enforce standards from collectively enforced covenants like the FFMC, or both.

My analysis shifts towards the practical dynamics within the Fediverse, where, contrary to the risks of authoritarian tendencies suggested by prior research, I observe a trend of cooperation and mutual regulation. My empirical evidence suggests a more nuanced understanding of decentralized community governance, where the absence of a central authority does not necessarily lead to toxic technocultures but may foster a distinct ecosystem of cooperation and collective accountability. The socio-informational ecosystem on the decentralized community network of the Fediverse, much like Reddit, is sustained by communities cooperating with each other and maintaining ties. On Reddit, cooperation can ultimately be enforced onto communities by a higher platform authority that makes it costly to break site wide guidelines (i.e., by removing the community from the platform). On the Fediverse, in the absence of a higher platform authority is a horizontal force of collective social sanctioning. That is, agreeing to federate with another instance creates an information channel for exchanging content and audiences across communities. Such social arrangements require trust between actors to sustain these exchanges and achieve their interests, and the possi-

bility of collective sanctioning has to be imposed to make it costly for one actor to impose a negative externality on another (Coleman, 1988). The existence of robust blocklists that are quickly adopted and shared across the Fediverse make it extremely costly to not cooperate or harm other communities (Colglazier et al., 2024). In the absence of a central platform authority, volunteer leaders across the Fediverse managed to collectively create a decentralized institutional force regulating the behaviors of communities by using defederation as an extremely high cost for compliance.

5.6 Conclusion

In this chapter, I sought out to explain why leaders of online communities choose to adopt similar rules of other online communities. For some, the reason was because it was simply cheaper to copy other rules that satisfied the initial problems of moderating content and regulating member behavior. For others, especially on the Fediverse, leaders copied the rules of other communities to address perceptions of legitimacy to external stakeholders, like newcomers and other community leaders. My analysis of interviews also revealed explanations for why communities may choose to adopt rules that embolden their differences rather than similarities with other communities. Rules may be adopted to reinforce the community's identity in the face of influxes of newcomers and overlapping membership with similar communities, helping to establish their niche space in a community ecosystem. In the next chapter, I describe how leaders understood their rules in the practice of community governance and the challenges and strategies associated with successfully achieving rule compliance.

Chapter 6

ENACTING COMMUNITY RULES: COMPLIANCE, ENFORCEMENT, AND TRUST

This chapter explores the multifaceted nature of online community governance in action. This is a departure from the previous chapter, which explored the motivations and choices in the design of community governance through rule adoption and rule writing. Governance in action involved the proactive and reactive strategies for rule compliance, including the enforcement of rules through punitive sanctions. The process of applying rules to problems in the community was revealed to be quite often complex as leaders felt constrained in their actions based on how community members might respond to being sanctioned for violating rules. Through a desire to enforce rules fairly and consistently, the actual meaning and the scope of rules would emerge when leaders had to decide on exactly when and where and how rules should be applied collectively in their teams. Through this practice of governance, leaders realized their decisions may make or break the credibility that gives their actions legitimate authority. Additional strategies were taken by leaders to foster trust and portray a level of fairness and consistency to not only make governance actions legitimate but also helped encourage community members to participate in improving the community themselves, leading to member retention and higher engagement. Finally, I synthesized these findings in the context of existing theories related to quasi-voluntary compliance, exit and voice in deteriorating firms, and the power of open communication in organizations.

6.1 Challenges for rule compliance: reading the rules

Compliance with the rules are crucial for the success and order of online communities. While the action to formalize norms into written rules is a helpful step in ensuring compliance within online communities, their adoption does not guarantee adherence. In fact, if the written rules of a community, made visible, ensured complete compliance and socialization to community norms, there would be little need for content moderators and administrators

to enforce rules at all. This is obviously not the case. Most leaders we spoke to across all the communities in this study described having to enforce rules and the challenges that enforcement brings to their volunteer work. For many leaders, this was a result of newcomers and members alike not even knowing the rules are in their community. A problem mentioned across all three settings in this study was the perception that people don't read the rules. FV13 explained how it's similar to the issue of people not reading the policies of websites they sign up for:

“I think, generally, it's accepted by society...that no one actually reads the terms and conditions. If you're signing up for Facebook or you're signing up to Twitter, you have no choice: you have to agree to their terms and conditions. You're not going to read them because they're incredibly long. You don't understand them because they're in legalese.”

Indeed, past scholarship has shown how and why people usually don't read the terms of service when signing up for apps and website (Fiesler et al., 2016). R7 had another explanation for why people may not read the rules: “I just don't think people take time to read the rules. I just think everyone thinks that they're not going to break any rules, so they just—they don't.” From this explanation, R7 implies that people come to online communities believing however they usually act is already appropriate or that they're not the kind of person who would intentionally break the rules. But, like FA3 points out, “people make mistakes—when they even know the rules”, and community leaders are tasked with teaching these rules to their members at any opportunity. Another leader, FA9, added more nuance to R7's point about why newcomers may not read the rules: “They don't read the guidelines. They assume that the guidelines are the guidelines at the Marvel Wiki or MyHeroAcademia wiki or whatever wiki they're coming from applies here.” In this explanation, newcomers arrive with their own set of beliefs about how to act based on other communities they've participated in on the Fandom platform. However, those sets of practices and beliefs don't always align with the new community in which these newcomers find themselves, and the result is leaders having to find ways to get newcomers to comply with community rules.

6.2 *Proactive strategies for rule compliance*

Leaders employed a few strategies to socialize new members with the rules. The first strategy involved a welcome message to new members who recently joined the community. These messages are usually pretty straightforward, as FA9 explains “We welcome new guys. We say ‘Welcome to the wiki, please read [the guidelines] because ours differ a little bit from others.” Some welcome messages are more detailed, such as FA11’s: “We tell them ‘Hi, welcome’ and then after that they will also be highly suggested that they read the discussions rules...and then we will tell the users about the rules, that we’ve updated them, if they could please take a look.” FV8 not only sends a message about the rules but also tries to help newcomers understand the platform in which the community exists and why it’s important to follow the rules:

“When somebody joins, I send them a big long email saying, ‘Okay, this is how Mastodon works. This is what the community is, here are all of the weird things that you’re not used to because they are special Mastodon things. It’s not the same as Twitter.’ Just to kind of help people get into it. That’s why I think that the rules are important for...a federated network, which is funded by the people who really use it, rather than with a company with lawyers.”

FV8’s more personalized and in depth welcome message is an attempt to not only socialize newcomers into the community’s rules but to also explain the purpose behind it in the first place and help them learn how to participate in what might be a different sociotechnical environment compared to what they’re used to.

Another strategy some leaders deployed was requiring newcomers to confirm that they have read and will acknowledge the rules before they’re allowed into the community. This was a requirement across several Fediverse communities, where joining a community often involves registering through an application form. For FV12, it was mostly a suggestion in their application, where FV12 said:

“It says ‘Read our Code of Conduct first, then get in touch if you’re interested in joining [our community]’. So it’s a very clear message that you should read

our code of conduct. We have a code of conduct, and we will most probably enforce it and follow it. And if you agree with it, you can join us.”

FV12’s concerns are with newcomers figuring out if their beliefs align with the community’s norms and goals, and reading the rules is a preview for what kind of community they’ll be in and how they’ll be moderated by the admins. FV13 included an attention check into the registration process to have a better understanding about who is actually reading the rules when signing up to become a member:

“And the last [rule], number seven: ‘When applying for an account, if you’ve actually read the rules, please enter “hippo cornflakes” into your reasons for joining.’ It’s sort of my canary to see whether people have actually read the rules. Most of the time, they don’t.”

For FV13, the attention check acts as the metaphorical “canary in the coal mine”, a signal to FV13 that the applicant has actually at least taken the time to read all the rules and even show willingness to comply before becoming a member.

Finally, another strategy for getting newcomers and members alike to be aware of the rules, including rule changes, was to use asynchronous communication channels or parts of the website to list the rules or announce rule changes. FV7 used an “announcements section pretty regularly” to post new or changed rules for their community. In communities like R7’s, information about rules is made visible at every opportunity: “We put up warnings and we pin posts with examples with reminders on breaking the rules.” R9 also made announcement posts in their community but admitted “very few people end up seeing our Reddit announcements, so people tend to be kind of surprised by rule changes.” R9’s experience was similar to most leaders we spoke with. Despite deploying all of these proactive strategies to get newcomers and members to read the rules, many do not. Instead, socialization may occur reactively through rule enforcement—when members break the rules.

6.3 *Enforcing the rules*

While proactive strategies aim to prevent rule violations, enforcement remains the most important strategy for handling rule offenses. Rules will eventually be broken. Whether a rule is broken because a new community member forgot what the rules are, hasn't read the rule list, or intentionally wanted to cause conflict, leaders of online communities at some point have to enforce the rules they've adopted to help ensure community goals are met and participants maintain civility. Although proactive strategies like directing newcomers to the written rules or having them confirm they've read the rules before they can participate are a good first step toward compliance and socialization into the community, most leaders we spoke to frequently encountered situations where they had to enforce rules reactively. I observed some patterns in the rule enforcement process across interviews. First, leaders rely heavily on their community members to help with rule enforcement and signal potential rule violation cases, usually through report mechanisms embedded in the website where the community is located. Leaders also varied in the strictness in which they applied their rules to situations, in what I call "hard" vs. "soft" rule enforcement. Finally, many leaders we spoke to described the challenges of managing the emotional labor of their work, usually as a result of a rule violator retaliating when being called out or sanctioned for their behavior.

6.3.1 *Relying on members*

As online community leaders are essentially volunteers, they're limited in their capacity to monitor every interaction and contribution in their communities. While they ultimately have the authority to enact governance by enforcing rules and applying sanctions to deter others from violating them, leaders relied heavily on their community members. FA12 explained this dynamic in how their community handles governing inappropriate behaviors:

"We have to be like pretty vigilant toward those sorts of behaviors, which is kind of difficult when our moderator team is so small. But it definitely helps that like other users have been stepping up and, you know, letting us know if one person is saying this thing and that it's a violation of our rules."

Members like these are interested in their community succeeding and help the moderator team scale their work by discovering and reporting problems that the moderator team may have otherwise missed because of their limited capacity. As FA11 explained “the users that are really, really active” will help enforce some of the rules as well. Fandom communities especially allow regular members to enforce rules on others, like correcting formatting issues in contributions to their wiki or reverting articles that have been vandalized. FA4 explained that “a lot of the community is like pretty good at kind of self-governing—self-correcting itself,” to refer to the outcome of affording members the authority to enforce rules at this level. For FA14, affording a low level of authority to help enforce rules was seen as effective governance: “It’s important to rely on your ordinary users—to have to give them trust and to let them enforce a bit of the rules themselves...because it’s very effective, honestly.”

The relationship between leaders and members on subreddit and Fediverse communities varied from those on Fandom. In both sites, administrative permissions are granted to those selected by the head administrator or moderator of the community, and everyone else in the community is subjected to their authority but may communicate instances of rule violations to their leaders. R4 explained that their moderator team “[relies] on the community a lot to help us moderate,” but this help almost exclusively occurs through reporting rule breaking content or behaviors. R8 explained how its members reporting rule violations helps the moderator team become aware of situations at all: “Them flagging stuff is what causes stuff to be seen, and then they definitely play a large role there.” Similarly, leaders of Fediverse communities expressed similar opinions about member participation in governance. When referring to how they enforce rules, most Fediverse leaders mentioned that they were responding to reports. Built into the sociotechnical systems for both Reddit and most Fediverse communities (especially those using Mastodon) are report and moderation queue features that help automate this process of signaling rule violations to authority figures who can investigate and potentially enforce rules. FV2 explained how rule enforcement occurs through this process:

“First and foremost, it’s reacting to reports, right? So there’s a reporting function in Mastodon, which the users make use of, and they see something offending.

And then that report pops up in the moderation interface for the administrator or moderator. And then you investigate what happened there, look at the content, and then figure out why the user reported it. And then you figure out, okay, does that really —is the reason to use a report based just on the subjective opinion of the user? Or is it really offending the rules of the server, of the instance? And then if it is offending, then you basically either remove the content or remove the user account that posted the content.”

The process FV2 described was echoed in other interviews with Fediverse and subreddit leaders. However, as FV2 demonstrates, not every reported rule violation provides a clear context for leaders to understand how to enforce the community rules.

6.3.2 *Graduated sanctions*

Leaders differed in how they approached rule violations, although a common pattern that emerged in interviews was graduated sanctions. First time rule violators may be messaged with a warning about their behavior, and only over time do the sanctions increase in severity as first time violators become repeat offenders. R4 explained how this approach to moderation differed from other leaders on Reddit when they said “other moderators will be very quick to jump to the ban, whereas I always scale up to that.” Allowing for first time offenders to learn about community norms with a light sanction and warning may help reinforce them as sanctions proportionate to the offense may be perceived as more legitimate and fair(Ostrom, 1990; Kraut et al., 2012). This practice also allows members to “save face” by giving them an opportunity to explain their actions and to feel like the sanctioning was deserved because it at least goes through a fair procedure(Ostrom, 1990; Kraut et al., 2012). While obvious bad actors such as trolls or vandals may be met with immediate severe sanctions to limit the damage they can do to an online community, leaders often regarded their good faith participants as capable of making and learning from mistakes. The graduated sanctions model is encouraged by sociotechnical system theorists because it can help enhance compliance to group norms and facilitate online community success(Kraut et al., 2012).

FV13 described this process on the Fediverse: “If it’s on my server, then I can have that conversation with them. I can send them that direct message. I can say, ‘Hey, these things you said on these Toots were not okay, you know? Stop it.’”

6.3.3 Deviations from the graduated sanctions model

Despite the advantages a graduated sanctions model for community governance entails, I find in these interviews cases where leaders deviated. Of course leaders recognize bad actors like trolls, vandals, and bots and have strategies to limit their damage with instant removals and bans. That said, leaders we spoke to justified skipping to severe sanctions rather than escalating sanctions for several different scenarios. First, some leaders found that responding to rule violations that disrupt delicate norms in the community was a useful deterrent for unwanted behaviors and limited the damage that breaking those norms can cause. Second, some leaders felt that affiliations with members or communities associated with hate groups threatened the safety and identity of their own community. Immediately removing those affiliations without any opportunity to save face was a common practice, specifically for Fediverse communities. Finally, it’s also possible that leaders may immediately issue a severe sanction and not give rule violators an opportunity to save face not because they’re inherently bad actors but because the voluntary nature of community governance, especially for large communities, leaves some with little time and energy to process violations on a case-by-case basis.

Reinforcing delicate norms

Leaders we spoke to described different scenarios in which the violation of community norms, whether or not in good or bad faith, could be so disruptive that a graduated sanctions approach would not suffice. Instead, these leaders would jump immediately to severe sanctions. For some communities, this was related to norms about the kind of content that was allowed. For example, FA4 explained how their wiki’s image policy is strictly enforced:

“We have a pretty relatively strict image policy on how images need to be named and detailed into our site. It’s a bit of work to get editors to follow them, though.

So we have to sometimes temporarily block people to try to, well, in a certain way, get them to understand, so to speak, that they need their edits to be up to our standards.”

FA4 believed that the only way to get contributors to follow their high standards for editing their wiki with image uploads was to block them from being able to edit at all if they violate the image policy guidelines. A graduated sanctions approach would involve removing or reverting the edit and communicating to the rule violator what the appropriate way to upload images would be in hopes that the contributor would apply that knowledge to future contributions. FA4’s governance team may have tried this in the past but it appears that contributors were not effectively learning about how serious the rule violation actually was. The result is communicating the severity of the violation with an equally severe sanction: the temporary blocking from being able to participate.

Another deviation from the graduated sanctions approach occurred because of medical misinformation that propagated in online communities on Reddit. For R5, members sharing misinformation about treatments during the COVID-19 pandemic were handled with little procedural fairness. R5 explained:

“You know, we have tons of information going on about COVID. But folks that would come in and talk ‘ivermectin’—talking things that have obviously been proven fake: deleted. And if you fucking keep it up, I’m gonna ban you. I’m gonna give you a couple warnings, but I have no patience for that, and you get banned. Have a nice day.”

In this example, R5 said they would immediately remove content related to unsupported treatments for the coronavirus like ivermectin, and that repeatedly sharing this content will result in an immediate ban from their community. Ivermectin is an antiparasitic drug that gained attention during the COVID-19 pandemic as an alleged treatment for the virus, despite there being no scientific evidence supporting its efficacy. During the COVID-19 pandemic, misinformation about treatments for the then under researched virus proliferated across social media networks, and because there was no evidence supporting it, leaders

like R5 were worried that their people reading those posts may believe those are viable treatments for a deadly disease. R5 remarked that they “have no patience for that” kind of content being shared in their community, and this sentiment framed their approach in how to handle it—severe sanctioning of the offender.

Perhaps more widely discussed in interviews was the way in which community leaders responded to instances of incivility between members. While a graduated sanctions approach may be used to encourage compliance with community norms, some leaders skipped to severe sanctions when behaviors became too extreme and antisocial. For example, R6 explained how they deviate from graduated sanctions if arguments in their subreddit’s discussions involve ad hominem attacks on others:

“The civility [rule] depends on how uncivil you’re being. If tempers are getting heated, it might just be everybody needs to calm the hell down. If it descends into...ad hominem attacks, the hammer will come up.”

In this example, R6 demonstrates first that the rule related to incivility in their community is ambiguous. Incivility seems to exist on a spectrum, from “heated tempers” to “ad hominem attacks.” Conflict is likely to occur between members of any online community, but when it turns to personal attacks between members, these behaviors can ruin the experience for everyone or if unchecked make it appear as if personal attacks are the norm (Kraut et al., 2012). R6 opts out of light sanctioning if someone engages in ad hominem attacks on other community members, responding with “the hammer” (i.e., a ban). R5 described a similar caveat to a graduated sanctions approach to incivility:

“What I usually try to do is, if it’s obviously a conversation that’s heating up and they’re getting amped up, usually a one day ban. If they have a repeat behavior of it...then we will do [permanent] bans.”

In this example, however, the graduated sanctions *begins* with a one day ban and only escalates to a permanent ban from the community. Repeat offenders of incivility rules like this threaten peaceful discussions in communities, and even if they’re not explicitly trolling,

their behaviors simply don't align with norms on how civil discussions should be carried out.

R3 justified a severe approach to antisocial behavior in their community because they felt like it was the best way to minimize the damage those behaviors can cause to the community. R3 said:

“And I think that part of the job as a moderator is to see that behavior and remove it from the equation before it gets out of control. And I reiterate—that’s the motion, being a moderator is stopping things before they get out of control.”

For R3, a hard approach to incivility in their community is “part of the job” because it leaves little room for those behaviors to fester and spill out into future discussions. Additionally, this approach lets leaders like R3 feel like they’re maintaining total control over the civility of interactions in their community.

In FA5’s community, the way in which certain words are used determines whether leaders deviate from a graduated sanctions model. FA5 described a situation where the use of a particular slur in a general rather than targeted application, while not permitted, would be responded to with only a warning. On the other hand, a targeted use of the slur would result in more severe sanctions. FA5 explained:

“We try to keep [sanctions] based on severity, where if someone says the R-slur, like the ablest one, but they’re not saying it like particularly to a specific person, they’re just saying like ‘this idea is [R-slur]’, we try to keep it more like a warning to where it’s just like ‘We don’t want to see that, but you’re not really attacking people.’ We have kind of the escalation where going after someone with some kind of slur, like the R-word—immediate consequence, which usually results in a ban because that’s not accepted here.”

The immediate consequence of a ban for using a slur, targeted at another member, is a deviation from the graduated sanctions approach and doesn’t allow for an opportunity to save face. It also sends a message not only to the rule offender but others observing the situation that behaving in that way is “not accepted here.” R2 described a similar reason

for immediately applying severe sanctions when hate speech was used during a discussion in their community. R2 explained: “The reason we shut down that discussion is because we can’t allow these bigoted views to proliferate. When you allow that, they just spread, they encourage.” Echoing earlier beliefs about stopping behaviors before they spill over into other interactions, R2 believes that a hard approach is necessary because it limits the damage those kinds of behaviors can cause to the community.

Limiting affiliations with hate groups

A problem for some leaders (specifically those on the Fediverse) was how their affiliation with other members of other communities impacted their own community’s safety and identity. Fediverse leaders described their governance work in terms of monitoring incoming content from other instances in which they were federate, or as FV15 called it, “taming the firehose.” Because anyone on their instance can see the content coming in from fediverse instances with which theirs was federated, members might get exposed to unwanted or even harmful content. One particular issue was content associated with hate groups. Fediverse leaders we spoke to described no soft or graduated sanctions model for approaching hate content and even users associated with hate groups. FV7 described one instance where this occurred and how they dealt with it:

“One [user] from two months ago made some pretty clear posts that had Nazi imagery—and there’s no other way to say that! That’s an immediate block, right? There’s no way I can allow something like that to come through and not block that.”

That FV7’s response is an “immediate block” shows how little tolerance some communities are toward content associated with hateful ideologies and hate groups. In fact, Nazi imagery was a recurrent problem for Fediverse leaders. So much so that that a trend among some leaders was communicating with other Fediverse admins which instances out there are associated with Nazis so that they can be preemptively blocked. FV3 explained this approach:

“So as far as defederation stuff goes, somebody will toot¹ that, for the sake of argument, ‘Nazi.com appears to be a Nazi site. We suggest you defederate.’ Okay, great, so I go and have a quick check, and, yep, sure enough, Nazi.com looks like Nazis. So you defederate. That’s sort of easy.”

In FV3’s example, an admin will warn them of an instance associated with a hate group, FV3 will then gather evidence to confirm the admin’s warning, and they will then make the decision to defederate (i.e., block) that Fediverse instance and its members from being able to interact with FV3’s. These warnings were often communicated across the social networks on the Fediverse with the hashtag #fediblock. FV11 described the strategy in more depth:

“I don’t know if you know about the #fediblock tag and stuff like that, but we instance admins kind of communicate with each other, warn each other about bad instances. It’s usually checking that, following up, doing a bit of research...So it’s just a tag with people. If they find an instance that’s got lots of racist people or something on it, they’ll post and put that tag on.”

FV11 and others would even find lists of known instances related to hate groups and share them with other leaders to preemptively defederate.

Defederation is deemed a “last resort” option in guidelines governing relations between Fediverse instances, like those in the Friendly Fediverse Moderation Covenant. Nevertheless, leaders we spoke to would decide to defederate without giving offending instances a chance to save face. FV4 justified such decisions by evoking the technolibertarian premise of the Fediverse: “I’m totally okay with blocking an instance and just saying we don’t want to communicate with them. And again, we’re not stopping their ability to communicate, we just don’t want to talk to them. They can go talk among themselves.” In other words, Fediverse leaders believe that other instances do not have a right to interact with their own instances, and that defederation is not the same as censoring or limiting the offending instances ability to free speech. Escalating immediately to this level of severity in sanctions,

¹On Fediverse platforms like Mastodon, the act of sharing content in a microblog post is common referred to as a “toot.” This is a reference to the sound that the elephant like creatures the platform is named after might make through their trunk.

despite being looked down upon and somewhat contradicting more widely shared norms in the Fediverse, is understood as fair because it also aligns with widely shared norms that hate groups aren't entitled to be federated with and because this action doesn't limit what those on the offending Fediverse instance can say.

Managing the labor costs of community moderation

I found one other scenario where the graduated sanctions model of community governance was avoided even when rule violators weren't bad actors. For R8, the problem of doing graduated sanctions was that it created more work than their staff of volunteer moderators can realistically manage. This problem was exacerbated by their community's size and popularity. R8 explained:

“In a subreddit that's very big, [it] has a lot of rule-type stuff that happens every day, and it also becomes a big drain on energy to just deal with everyone on a case-by-case basis. And so lots of things that might be fine elsewhere aren't given a second chance. Bans will happen often for things that could be fine elsewhere.”

R8 acknowledges that other communities might have a graduated sanctions approach to rule violations, but on their community, people who violate rules aren't given any opportunity to save face or a second chance at participating and are instead banned immediately. While this may seem harsh, the reality of volunteer governance for popular, large communities like R8's is that there are too many people to manage. In other communities, leaders will remove the offending content and kindly message the offending user with a warning about the rule they've broken. For R8's community, doing this for each and every rule violation “becomes a big drain on energy.” Leaders of subreddit communities are volunteers who typically govern their communities when they can, but as R8 demonstrates, when that work is scaled up because of the community's size, it may not be feasible to have such a hands on approach, where rule offenders are given an opportunity to plead their case and given a second chance to learn from their mistakes.

6.3.4 *Managing emotional labor*

One of the biggest challenges that leaders mentioned in interviews in regards to enforcing rules is the emotional labor it frequently entails. By enforcing a rule on a participant of an online community, leaders go from bystanders to inserting themselves into a problematic situation and dealing with someone who might already be upset or particularly volatile to being told their actions are inappropriate. R1 described this aspect of governance as “the hardest part” of their work:

“...because it very often comes down to being a personal thing for a poster. Every once in a while we’ll say like, ‘Hey, you’re not allowed to post this because what you’re saying is essentially racist.’ Of course, that will set the poster off, because [they’ll respond] ‘I’m not racist. How can I be racist?’”

What typically follows is a series of exchanges between the rule offender and the moderator of the community that’s addressing the problem. For FV9, they noticed that “people who break community rules like to quibble with whether or not what they did was a specific violation of a specific rule,” and that they ultimately “don’t want to get into that.” R8 said that such exchanges “could be very draining...just the constant rehashing of it over and over again.” Overall, people who violate community rules and are called out and sanctioned for it may create additional burdens for volunteer leaders, who must then manage arguments and justify their decisions. This work isn’t necessarily easy either. FA14 said that when enforcing rules and issuing sanctions, “you have to keep your calm and manage the situation, [but] sometimes you don’t keep your calm. And it’s quite hard to be a perfect moderator in some ways.” Leaders strive to remain neutral arbiters of their community’s rules, but the way that rule offenders respond to being sanctioned can be unpredictable, and sometimes volatile.

Many leaders we spoke to did indeed encounter volatile and retaliatory behavior from people who were sanctioned for violating rules. Some responses are relatively mild, like when FV9 had to enforce a rule on a member of their community and “they kind of threw a fit” and argued. Others, as R5 described, may escalate quickly because some people “just

fucking hate your guts because you deleted something they said one time.” Enforcing a rule on an individual may even bring out other members who have harbored grudges or resentment about the community’s leadership. R5 explained:

“If you do something that’s unpopular, you will get the absolute shit dogpile badly, and people will come out of the woodwork to take the opportunity to vent at you particularly. And that’s something I keep in mind with moderating.”

The threat of being verbally abused for enforcing a rule meant leaders had to be cautious and make decisions carefully. R3 explained how this dilemma might mean not enforcing a rule at all:

“So that’s kind of the linchpin for us is that if something gets posted that technically would violate the rules but if it’s got like 500 upvotes and there’s like 900 comments on it...we’ll keep it up there. We won’t remove it.”

This example revealed the balancing act that leaders manage between wanting to uphold rules but also maintain goodwill with their community members. By removing a popular post that’s received a lot of activity from their community members, they risk backlash and the potential for conflict by disrupting the community’s discussions.

Leaders mentioned options for managing the emotional labor that accompanies rule enforcement. Some took proactive measures against members that appeared to be participating in bad faith from their entry into the community. For FA3 this meant blocking the user before they got out of hand. FA3 explained their reasoning saying, “If it’s something where it’s like ‘No, this person is just here to troll us,’ then we just instantly block them.” For problematic Fediverse instances, FV13 said they would “often either silence or block the entire remote server because it’s probably not worth the effort,” after they discovered that the offending instance lacked moderation for the kind of behavior that isn’t allowed in FV13’s community. For R6, they’d only temporarily ban someone for breaking a rule “unless the user then throws an absolute shit fit at the moderators, in which case, the hammer falls...we’re done.” In all of these instances, leaders manage the emotional labor of dealing with particularly volatile or disruptive participants with a zero tolerance “hard” approach,

removing the problem makers with no room for debate. In other words, emotional labor is managed by not trying to argue and teach the rule violator why their behaviors aren't welcomed, as it will likely not be worth the leaders' time to do so. R7's moderation team had to take additional measures to protect their emotional well being because the backlash they would receive from enforcing rules turned into greater threats of targeted harassment and defamation. Some of their leaders would have to create alternative Reddit accounts to help maintain their anonymity, so that they can have a separate account for participating in other communities on Reddit. R7 explained:

“The reason why some of us have backup accounts is because when we're enforcing our rules on our community, some people just get upset and start reporting our usernames and stuff or harass me. I got one [report] for sexual harassment a couple weeks ago. I haven't even sexually harassed anybody! And while [Reddit's admins] are figuring it out, sometimes it results in a temporary ban.”

Simply performing their duties to govern their community can lead to the kind of harassment that can get R7's account banned from Reddit's entire platform. Upset rule offenders will try to use Reddit's reporting mechanism to retaliate against the actions they take by enforcing rules.

The challenges leaders faced and strategies they deployed illustrate that rule enforcement in online communities is far from a straightforward task. It involves navigating not only community reactions to rule enforcement but also the nuanced and complex nature of the written rules themselves. The complexity leaders described with respect to how they enforce rules highlights the fact that written rules frequently require interpretation. Despite the intricacies of their wording, the written rules aren't always effective as a map for navigating ambiguous situations, challenging leaders to continually reinterpret and refine their understanding and application of the rules to ensure fair and effective governance over their communities.

6.4 *Socially constructing rule meaning*

Effective enforcement often requires nuanced interpretation of rules. Deeper analysis into interviews revealed that in teams with more than one leader, leaders encountering ambiguous moderation situations often consult their peers. Together, they interpret the situation and its relation to community rules, continuously redefining or reinforcing these rules through their enforcement. This process demonstrates a variation of “interpretive flexibility,” (Kline and Pinch, 1996) as leaders, shaped by their personal experiences and the context of their community, may or may not apply the same rule in different ways. Leaders as well as members may come to a sociotechnical system with prior beliefs about how people *should* interact (Orlikowski and Gash, 1994), and those beliefs may be misaligned with the community’s goals. What I observed from interviews was leaders collaborating on defining a consistent way to interpret the meaning of interacting in these social systems and how the rules must be applied to maintain community goals. In other words, this collective process not only helps build a nuanced understanding of the rules, either remembered by veteran moderators or documented for future reference, but also ensures consistency in rule application. Such consistency is crucial as it is perceived by community members as a form of fair governance, reducing uncertainty about expected behaviors. Fair and consistent rule enforcement strengthens trust with the community members, which builds credibility toward the legitimacy of governance actions. When members perceive enforcement as legitimate, they are more likely to adhere to rules and even assist in enforcing them, like reporting violations or correcting newcomers’ actions.

Assessing the gray area

Despite the efforts in strategically selecting rules for the community as well as trying to get members to read the written rules, leaders we spoke to encountered ambiguous situations where it wasn’t clear if a rule had been violated. For some this was related to interpreting behavior rather than content. R1 for example described how the decision to moderate the way someone posts is like “the biggest gray area...the biggest headache because that’s the part where it’s hard to regulate by clear rules.” Indeed, non-verbal communication

in computer-mediated settings may result in misunderstandings as context and knowledge about the speaker is often missing (Walther, 1996). Rules that restrict particular kinds of behavior rely on not only the moderator's interpretation of the content of a user's speech but also an assumption of their motives, intent, or charity toward the situation. R5 had similar challenges with the "gray area" of moderating for their community's "Don't be a jerk" rule, of which they said "sometimes the scale of don't be a jerk can vary." In some cases they let the community decide by using Reddit's downvote feature to push content down to the bottom of the page, but in others they may subjectively interpret some situations as too "egregious" and quickly sanction the user by deleting their content before it disrupts community members.

FV10 had challenges with how to respond in situations where a user was accused of violating their community's rule on harassment. For FV10, this "gray area" comes down to what the appropriate response should be: "Do we try to have a conversation with them to get them to sort of tone it down versus do we strictly apply this rule as a basis for removing their access?" FV10 resolved this by making the definition in the harassment rule "squishy...on purpose, so that there can be some flexibility in terms of how rigidly to enforce it and how much tolerance we're gonna have for sort of misbehavior or people kind of getting it wrong." Leaders may encounter entirely new kinds of situations where a certain type of content may or may not be violating a community rule. FA6 experienced this when trying to enforce a rule barring the sharing of "speculative" information related to news about the topic of their Fandom community. In the past, this was easy to moderate because most sources that contributors would use for rumors were almost always from unaffiliated or anonymous entities. However they encountered a new situation where a rumor was backed by an official source affiliated with the game the Fandom is based on. Even when rules had definitions tied to prohibited behavior, leaders we spoke to still described problems in how they should be interpreted. For example, FA12 explained here how interpreting behavioral rules like one against "discriminatory comments" presents a dilemma:

"I think our definition of like making discriminatory comments, or something like that was a bit up to like, you know, our interpretation...It's still honestly a

bit muddled because, like, we have to look at it and be like, Well, is this really breaking the rules? Or like, it's hard because like, obviously, if someone is being discriminatory, obviously, you wouldn't want them in this community because it's supposed to be like, more safe and welcoming. But on the other hand, it's like, does this actually—like, you know, obviously, what this person is saying might not be popular, but is it actually like a violation of this and this rule?"

FA12 highlights the challenge in enforcing behavioral rules as they frequently require subjective interpretation. Despite having rules against “discriminatory comments,” the leaders find themselves in a dilemma over distinguishing between unpopular opinions and genuinely harmful statements that breach community standards. When the situation isn't clear, leaders had to engage in an interpretive process, often in collaboration with other leaders, that simultaneously constructs a new meaning of the rule.

Interpreting rules collaboratively

The process for interpreting ambiguous situations as cases of rule violations or not often involves a collaboration between the primary investigator of the situation and their co-leaders (co-moderators, co-admins, etc.) R9 described these situations as “gray areas that we then have to talk about as a mod team, what side of the line will it end up falling on.” New leaders may have difficulties in knowing if their interpretation of a situation constitutes a rule violation and they may seek feedback from more experienced leaders on their governance team. R3 described how this process played out when they were a new moderator of their community:

“So there's been times where like, when I first started out, I'll be like, Okay, I would remove this post, but I don't know if another moderator would, so I would ask for a second opinion. So I would post a link to it, say, Hey, this is the story. What do you guys think? And they would comment and say, Yes, remove it and give reasoning why.” Other leaders like R5 would reflect on their own subjectivity and request feedback from someone on their governance team who wasn't as emotionally invested in the situation: “So we're, we're actually

connected on Facebook. So if I've got a situation where maybe I know I'm getting a little hot and emotional, I'll run it by him. And vice versa. So we kind of check each other, and consult each other when there's a situation that needs it."

R5 and other leaders we spoke to often used external communication channels through messages on Facebook or even private Discord servers for discussing issues in their communities as well as for helping troubleshoot situations that land in the gray area of rule violations. R6 echoed this strategy saying "If there's something that isn't easily handled, there's a separate facility for us to trade ideas." Indeed, in the situation mentioned earlier from FA6, the leaders of their Fandom held a discussion to decide how to handle speculative news that, while normally prohibited, may not be a clear rule violation if it comes from an official or affiliated source: "We sort of rationalize that if—as long as it's from [someone working at the company], it's confirmed." From here, FA6's team has produced a more nuanced understanding of their rule on speculative news by creating a caveat based on the circumstances under which such information can be deemed reliable and permissible.

Examples of collaborating on interpreting a gray area situation also extended to the decision to defederate from other instances on the Fediverse. FV8 encountered a complicated dilemma where an instance for sex workers connected to their community, resulting in their community's feed being populated with pornographic material. On one hand, FV8 recognized this instance as a safe space for sex workers to support each other and find community, but on the other hand, FV8 also wanted to protect their own community from seeing this kind of content on their feed. Unsure about how to handle the situation, another member of FV8's community pointed out a way to interpret one of their rules so that they could block the problematic instance: "And then finally, somebody pointed out to me that actually, in our server rules, there is a line—they were breaching the rules, because you're not allowed to do that. Essentially, it was non consensual, right? Non consensual sexual content. So I was like, Oh, great. Now I've got a legit reason I can block them." Legitimate governance actions were valued by leaders we spoke to because ultimately they want to be perceived as fairly enforcing rules. A similar example occurred for FV12 in

which permitted posts for abortion funds were being trolled by another community member who levied dissenting views about abortion and equated it to “murder”, making other community members upset. FV12 struggled to see how their current rule set could apply to this situation:

“But then I went to the Code of Conduct, looking where the Code of Conduct is –what is the argument that I can put forward? How can the Code of Conduct be applied to my idea that this is unacceptable? And I couldn’t find anything about abortion rights, because [the rules are] not so specific.”

FV12 had to resolve this issue somehow, and, unlike other admin and mod teams discussed earlier, they are the sole administrator of their instance on the Fediverse. Instead, they reexamined their rule set to see if it could be interpreted to fit their situation and their desired outcome (sanctioning the offensive member):

“But after a while, I realized no...if you say, a woman has no right to decide if they, if she can have an abortion, she is discriminated because she’s a woman. She’s discriminated in that sense. And saying that she should do a birth of a child, even if she doesn’t want to, or she was raped and she’s forced to give birth...this is discrimination. So it falls under this. So this was kind of like my dialogue with our own code of conduct.”

By opening up the rule against discrimination to this broader interpretation, FV12 was able to rationalize their decision to sanction this community member with a suspension. While not a collaboration with other leaders, FV12 engaged in a “dialogue” with their community’s rule text, ultimately reshaping the understanding of the discrimination rule to include reproductive rights. Overall, these examples illustrate how rule enforcement is an interpretative process in which rule meaning is socially constructed during ambiguous situations. Leaders may try to foresee all scenarios that disrupt community norms and adopt rules for them accordingly, but interviews show that this is not always the case. In fact many leaders we spoke to encountered edge cases or complex situations that didn’t align with the prohibited behavior outlined in the formal written rules but still disrupted

or caused conflict in the community. The persistence of those situations may lead to the revision of a written rule, usually through refining or expanding on the language within the rule text with definitions and examples. For R9, gray area situations where a decision was made to enforce a rule were frequently documented as “case rulings” in a shared Google document with other moderators. R9 explained that with the document, “it helps if we’re consistent with every single one of [the community members] for them to feel like we’re not favoring certain users over others.” It also served as a resource for other moderators in their community to model their responses for future ambiguous situations in similar contexts.

6.5 *Building legitimate authority*

Leaders are compelled to establish legitimate authority to ensure compliance with rules and maintain member retention, often by fostering fairness in how rules are enforced and building trust that leaders have member interests at heart (Levi, 1988; Levi and Sacks, 2009). Suchman (1995) defines legitimacy as “a generalized perception or assumption that the actions of an entity are desirable, proper, or appropriate within some socially constructed system of norms, values, beliefs, and definitions.” The perception of legitimacy helps to ensure that members view leaders’ actions as aligned with their communities’ values and expectations, which might make it easier for leaders to enforce rules without significant resistance. Like R9 mentions in the quote above, consistency in rule application is important for establishing trust between community members and leaders. Many leaders we spoke to stressed the importance of being perceived as fair when they enforce rules and went great lengths to establish trust with their community members to reinforce governance actions as legitimate. Some even opened up spaces within their community for members to voice concerns over problems they experience, give critiques on how leaders handle governance, or give feedback on proposed policy changes. Broadly, I find that leaders may be motivated to engage this way with their members to help them foster stronger bonds with their community, resulting in better retention as well as compliance with community rules. One way of understanding this relationship is through the lens of a Rousseauian “social contract”, in which individuals relinquish some of their freedoms in exchange for protection and an orderly society. Enforcing rules on members of online communities is often a restriction of

specific kinds of speech and behaviors but helps achieve the goal of the community (e.g., having contributions be relevant to its shared topic) and protects members from abuse. That said, an enormous amount of online communities exist, and creating a new one is certainly less costly than trying to create a new society or country when one feels like the way their online community leaders are unfair to its members or doing a poor job at governance. Retaining members when exit, while having some costs, is virtually an inexpensive option is an ongoing challenge for online community leaders. Through my analysis, I show how volunteer community leaders strategically enact governance to be perceived as fair and establish trust and credibility with their community members in order to avoid conflict and maintain community cohesion.

Portraying fairness

Across all three settings, leaders we spoke to expressed a desire to be perceived as fair to their members, especially when they enforced rules. One facet of fairness was expressed through the concept of “consistency.” FV1 baked consistency into their moderation philosophy: “I think...people appreciate a steady, even hand; a consistent tone or consistent approach” to rule enforcement. FA11 extended this idea of consistency of rule application: “I think the moderator team is like vigilant enough to apply them evenly everywhere to everyone—even their friends don’t get passes.” That not even the moderators friends “get passes” for violating rules implies that FA11’s moderator team tries to show they are as unbiased as possible, so that other members don’t feel like they’re being unfairly treated.

For FA12, consistency meant to be “as unbiased as possible” when enforcing the rules of their community. Echoing this sentiment, R3 said “we always try to maintain neutrality when we moderate” and that their fellow leaders strive to avoid acting acting as a “madmin”, which is a colloquial term (mad + admin) they used to describe leaders who “just hate everything that doesn’t agree with them and they will delete all your comments if you don’t agree with them.” Having experienced a madmin situation in the past, R3 understands how this totalitarian approach to governance alienates community members and makes it harder for them to trust in the legitimacy of their leaders’ actions. FA5 described a similar situation

with a leader in their community who ramped up the strictness of their interpretation of the rules on speech, causing members of the community to complain, saying “we can’t really converse in there anymore.” FA5 took these complaints seriously and tried to resolve the issue:

“We did pull [them] aside and were like ‘Hey, we got to work on how we’re gonna handle these rules because we’re getting these complaints. And at the end of the day, if the community is feeling kind of suffocated, then maybe we need to not be so hardline on handling these things.’”

FV17 reflected on their experience in adding more rules that restrict community members behaviors and speech and what consequences this has for the community:

“The more rules you’ve set up, the more people were pissed off, and there are people who come up with splinter groups and then there’s all kinds of really toxic behavior and it’s really ugly.”

Overall, leaders strive to avoid overreach and over-regulating because it may lead to unintended conflict in their community. R6 echoed this belief: “I’m not a huge believer in over-regulating because I think that ends up being counterproductive and produces outcomes that you weren’t actually intending.” For R3, “that is one of the hardest parts” because they deeply care about the topic that’s being discussed but are “just doing [their] job as a moderator” when members believe they’re overstepping in rule enforcement. It’s a fine line to balance between ensuring members feel comfortable enough to openly discuss the topic and enforcing rules when discussions become uncivil or veer too far off topic.

Finally, I noticed many leaders approach governance by, in most cases, assuming first that members who violate rules did so in good faith. FA3 described what this meant for their community: “So if they make a mistake, and it’s not ill-intentioned, we do not hold it against people. Like, people make mistakes when they even know the rules.” Leaders like FA3 take a soft approach to how they enforce rules and understand that everyone makes a mistake and may not be trying to harm the community. Other examples of assuming good faith were followed with leaders describing seeking out clarity about a situation from

the rule violator. FV7 described this process: “We try to get clarity...we go to them and say, you know, we need to understand what you meant here. Can you be clear on this?” As mentioned earlier, interpreting a rule violation situation can be challenging especially in these non-verbal, CMC settings. Seeking out additional context like this helps leaders reduce the ambiguity of the situation as well as give the offending community member an opportunity to explain their actions and intent. R5 described a similar process with a story about a member of their community who was reported for their behavior:

“So I had one person that was [reported]. I looked at their message history. They were going through a divorce. They had fallen off the wagon with addiction, and they just went fucking insane on the subreddit. Like, it was ridiculous, and it was nonstop. I put them in a timeout; I sent them a private message. I was like, ‘Hey, you’re going through a hard time right now, it’s just a one day timeout.’”

As someone who stated “I tend to know my users”, R5 demonstrated their capacity for understanding their community members as the complex, emotional human beings that they are. This understanding helps inform the way R5 enacts governance in their community, and in this example, they show that they communicate to the offending member that they understand what they’re going through and explain the sanction so that they don’t feel like they’re being treated unfairly. R1 reflected on the importance of this empathic approach to governing in their community:

“So that’s the kind of thing that feels like....as a moderator, we try to view almost all users in a good light until they prove us wrong, right? That’s because if you don’t do that, you’re not going to want to moderate anymore because you’re going to think humans are trash...so you try to think whatever prejudice they have is born out of ignorance, rather than, you know, malice.”

R1’s restraint in judging their members is important because it helps them regulate the emotional labor of dealing with conflict in their community. The other outcome of an empathetic approach is better compliance with the rules of the community. R8 stated “I like giving everyone as much chance as I can. I feel that most people...will usually be happy

to change their behavior with guidance.” In fact many leaders we spoke to echoed this sentiment that they would rather help someone learn how to behave in their community rather than have to punish them. That said, R8 also remarked on the challenge this presents when moderating communities with very large numbers of members: “So I would like to do that whenever I can, but when a subreddit is too big, there simply isn’t time to give that chance to everyone.” I reiterate that leaders we interviewed are volunteers who do this work with their own free time. In an ideal situation each rule offender would be given a chance to learn and redeem themselves, however, R8 has pointed out the limits of an empathetic approach to governance. To moderate at scale, many volunteer moderation teams utilize automated tools like AutoModerator for communicating rule infractions and sanctions to offending community members. While an empathetic governance approach helps foster rule adherence and mutual understanding, it may have its limits as communities grow in members, necessitating additional strategies for rule compliance.

6.5.1 Establishing trust with community members

In addition to striving for fairness and consistency in how rules are enforced, leaders we spoke to described other ways in which they build credibility as legitimate authority figures in their communities. Some of these include efforts to socialize newcomers by creating a safe, welcoming environment. Other approaches involved transparency and humility, like publicly acknowledging mistakes made in how they’ve been governing or problems that have deteriorated the quality of community discussions. Finally, some leaders opened up avenues for members to voice criticism about how governance is enacted or to engage with leaders on how to solve problems or vote on policy changes. In sum, I argue that these tactics help build trust in community leadership. Leaders who enforce rules fairly, communicate openly, and seek community feedback enhance the benefits of being governed by self-appointed authorities. While volunteer leaders can effectively become “benevolent dictators”, interviews suggest that approaches to governance in online communities vary. Incorporating various democratic methods into community governance are perhaps not even optional for some leaders. To gain the consent of those being governed, leaders may have to show and

practice that their decisions reflect the membership's collective will and interests.

Leaders actively engaged with newcomers and community members to help them feel safe and welcomed. For FA4, this was motivated from a place of empathy:

“I try to ensure that all users can feel welcome because I was at one point in their shoes. I remember that, for instance, something that I found that can be a little bit intimidating for a new user is, for instance, to have their contribution undone by another user. They can feel like maybe they do something wrong and that they are unwelcome. That's something that happened to me.”

Wikis are infamous for having trouble retaining newcomers because of how they react to having their contribution being reverted for not meeting the community's standards. FA4 took their experience with this as a newcomer to a wiki into consideration when deciding how they approach newcomers in their community:

“So what I like to do is explain and tell them that I appreciate that they want to contribute that, but I just give them a heads up and try to establish some sense of familiarity between any new users and the community.”

The “heads up” here is about the community's rules and how their contribution may have to be removed, but FA4 does this in a way that acknowledges the newcomers naivety of the community's rules and guides them toward better practices. In this way, newcomers may feel less like they're being attacked and more like someone in the community cares about them contributing in the right way. Other ways that leaders helped members feel welcome was by supporting their psychological safety. For example, FV13 required newcomers to sign up to this amateur radio community with their radio callsign as a way of filtering out people who aren't involved with the amateur radio community. However, this was a problem for one newcomer:

“We had someone contact me who wanted to avoid being doxxed because their call sign was registered in their deadname, so they didn't want to give away their callsign and therefore their dead name and all that information.”

FV13 agreed to let this newcomer bypass that registration requirement despite it being part of the community's rules. Leaders wanted their community members to feel like they can be safe and took effort to ensure this. One controversial event involving this occurred for FV16: "There was a case recently where somebody on one of the bigger free speech type instances where anything goes—they were being very aggressively transphobic to our members." FV16 blocked the user from being able to interact with FV16's instance entirely, which caused conflict among admins in both communities. Nevertheless, leaders sought this kind of safety for their members. For R1, it was integral in how they governed: "We try to make our community as safe as possible." R1's leaders "interpret negativity on a broad scale" when they moderate members' interactions and feel as if they "raise the bar a little higher" than how other communities on Reddit moderate things like personal attacks. R3 had similar beliefs about governing:

"I'm a moderator because I belong to this community, so I care about it. And I want to make sure that people aren't getting scammed, that people are coming into the subreddit and being able to ask a question without being attacked, or being called an idiot, or being called stupid. Some discussions can get out of control and we step in and just go 'Listen, we get it, but it's not the place.'"

Regulating speech and behavior in online communities to protect members from harassment and abuse is simultaneously a restricting governance action but also one that allows people to feel comfortable and safe enough to participate at all. At one point in FA12's community's history, its leaders were pushed out for being "lax with users who were being discriminatory, which was very disheartening to see for a while." They had a "big overhaul of the moderator team, and because of that, a lot of things have definitely changed." This occurred in R9's community before they joined its moderation team as the previous leader was "pushed out because they weren't putting in any rules to protect people from scams." In both of these examples, leaders that failed to protect members from harm were pushed out of their position and replaced with those who will. Despite the belief that leaders of online communities are inherently "benevolent dictators for life" (Schneider 2021), regimes may be overthrown when community members perceive that leaders have failed to fulfill their

obligations under the implicit social contract between them.

Opening channels of communication

One strategy leaders took to build even more trust with their community members was opening up channels of communication. Through communication channels, leaders solve problems, publicly acknowledge their mistakes, create spaces for feedback or criticism, and involve community members in discussions about policy changes. For example, members of R5's community experienced an issue where an automated moderation tool embedded in Reddit's sitewide moderation infrastructure was automatically deleting posts related to a fund to help women get abortions in their state. R5 described the situation and how they handled it through openly admitting to the issue and communicating to community members how they would fix it:

“So someone got thoroughly pissed off. They posted the Yellow Hammer Abortion Fund, where they will send you Plan B as well as whatever the drug name for the abortion pill is. Reddit auto-deleted it. But [community members] thought we did it. So they came back and they came back hard about how terrible moderation was. I don't delete that [post criticizing moderation]. I engaged with that user because that's a public thing, and I said, 'Hey, I'm looking into this. This was automatically deleted by Reddit. I'm gonna go over here [a forum for Reddit moderators], I'm gonna post the question and figure out what happened.' So I addressed the complaint publicly. I provided them with the research of what happened. I told them how I was able to fix it, and they could repost.”

R5 admitted to not knowing what was going on but also took the member's complaint to heart and did what they could to resolve the issue and find out why it happened in the first place. Left unchecked, this kind of criticism could fester into speculation about the unfairness or bias of the leaders of the community, which may cause members to trust the governance actions of its leaders less. Similarly, R1 had to write “an official apology for not monitoring Twitter” during an event where they had their members vote in a community

contest. Some members reposted the poll on Twitter, inviting lots of outsiders to skew the results. However, many were still upset, and to figure out how to best resolve the issue, R1's leadership team decided to ask the community members for guidance:

“So we offered—we did a poll in the community: do you guys want to do a revote or not? We [the leaders] weren't leaning towards a revote because it's a lot of work, but it was an offering. And then when we put that post out into the community for opinion, which included that we would invalidate the original results for [the contest name] because they [the nominee] won everything.”

R1's mod team eventually just randomly selected winners for the poll as the lack of resolution on the issue was only causing more conflict. The result was deemed satisfactory:

“So we posted those new results and it was amazing how our community goes, well, you know what, even though this was decided by a wheel of chance, it actually feels more fair and representative.”

It must be stated that going to these lengths to not only provide entertainment for their community members but to also put in this much effort to resolve conflicts created by it and satisfy community members is a lot of work for volunteer moderators. In the face of criticisms and complaints, R1 received mostly positive feedback to how they handled this conflict in their community:

“So like, it's that kind of positive feedback we get, [it] gives us a lot of confidence now for doing things more or less the right way, or the positive way. The way that builds a community that our most active users are interested in participating in, in the end. And although we do this, to a lot of people, the complainers, like we always tell them, you're welcome to build another community as you see fit. Because that's the beauty of Reddit. It's like one click starts a community. But building it up is hard.”

Exit is an option for disaffected community members, but in reality, it takes a lot of work not just in regulating speech and moderating content but also in negotiating with members

about how to resolve internal conflicts and create opportunities for creating the kinds of bonds and trust that build community and retain members.

Leaders of other communities opened up channels of communication in different ways. For example, some create spaces for criticism or feedback about how things are going on in the community. FA3 described one such example:

“We have this one which we call Wiki Matters where people just bring up their comments on anything going on in the wiki. So like right now we’re having a discussion on how to deal with overcategorization of pages.”

In FA3’s community, member input was valued even on rule discussions:

“If someone has an issue with the rule, they bring it up in the forum—the Discussions one—to an admin. And then we go through and we talk it out with them and the rest of the admins with the community at large, and figure out if it’s something reasonable that we can adapt to and address.”

In fact, leaders would actively seek out feedback about rules from their members. FA6 said that before they adopted rules in their community they “actually did outreach for a long time” to get feedback about what rules would be appropriate. In FA11’s community, “rules are added usually whenever the user base complains”, showing that policy changes are often driven by problems brought forth from community members who voice their concerns publicly in community forums.

In other cases, leaders involved community members more closely in governance and policy decisions. One such case was FV17, whose leadership team incorporated members during the creation of the community to help with describing the goals and creating the rules:

“We actually put it up as a public discussion in the instance itself. So we started blank, and we thought, you know, we wanted to involve everyone who is there with us from the beginning to have a say.”

For FV17, this approach was an essential part of their governing philosophy:

“At that point in time, I was really, really quite interested and quite idealistic in participatory communities. I was like, everyone should have a say, you know, we should use the technology to help us be fair.”

Here, FV17 equates participation in governance as an aspect of fairness. Participatory democracy is a strategy for having members be involved in decision making processes, and in turn, it helps members feel like their interests and concerns matter to the leaders. Such strategies may help minimize the risk of attrition as leaders create opportunities for members to establish stronger bonds with the community.

Another example involved incorporating members in almost every stage of policy discussions and decision making. R4 explained their philosophy to community participation in governance: “I think the good way to handle rule development is based off the community’s needs and consult them the whole way.” R4 frequently surveys their community members to understand any problems they’re having and how they wish the moderator team to resolve it. Unlike other communities where “moderators become sort of a force that directs the community to one thing”, R4 believes decisions about how the community is governed should come from the population of community members rather than directed by the few moderators who lead it. This belief translated into how R4 overhauled their community’s rules:

“To build those rules, we went through a process of essentially doing surveys every quarter for our community and saying ‘Hey, these are the problems that we’ve identified—does this resonate with you? What do you think we can do to fix this?’ We’ll take those suggestions, we’ll incorporate them or implement them, we’ll come back after a quarter and be like ‘Hey, we’ve done this because this helped.’”

R4 sees their role as a facilitator of policy change for their community members, to help them organize and collaborate on solutions to problems in the community that, as leaders, R4 and the other moderators can only draw speculations.

In fact, one particular example described by R4 showed how much influence their community’s members have in directing policy change. The community’s most active members

expressed frustration that their subreddit, intended as a forum for cybersecurity professionals to discuss industry news, was increasingly inundated with off-topic questions about personal cybersecurity and career advice, detracting from its goals and challenging the community's identity. R4 sent out a survey to its members acknowledging the problem and asking members how they feel about it and what the moderation team should do to resolve it. R4 said "the response from a lot of the regular, active members of the community who said they don't want any of this, if we keep seeing this, we might not come back." He added that many of the responses suggested "get rid of these, I'm here for [cybersecurity professional content]," referring to the surge in posts about personal cybersecurity and career advice questions. For R4, the threat of exit from its active members carried significant weight. R4 explained, "Those people will never get their questions answered if no one except for students [post here]. We need a way to have professionals to keep coming back, engaging in discussion." If the community is flooded with a cacophony of college students asking for career advice, people who want to actually discuss problems or news in the professional cybersecurity world will be drowned out and eventually just leave. Without those professionals around, the community would essentially fall apart. R4 ultimately created new rules forbidding personal cybersecurity and career advice questions in the community.

Some leaders we spoke to also polled their members in order to make a decision on rule change. R9 explained the process in their community:

"If we've seen people a little cranky about some rules that we have, we will put out a poll...to see if they want things to change or not...so we try to change things as the community wants things to be changed."

Before R9 joined the moderator team of their community, the previous moderators were pushed out for not doing enough to maintain their community and keep members satisfied. As a result, R9's moderator team tries to make sure they obtain the consent of their members before implementing a big change. This was reflected in how they interpret poll results:

"There's so many polls that end up basically 50/50, so we tend to leave them as is if they're 50/50. Or if it's like 51/49, we'll go with the 51 just because that's, like, out of our hands. We don't want to make that kind of decision."

By putting the decision-making responsibility in the hands of their community members, R9 is, like R4, facilitating governance change but not authoritatively directing it. In many of these examples, leaders seek out the consent (and often ideas) of those they govern when a problem in the community needs to be solved or if rules need to be created or changed. This reflects a variation in community governance that's more democratic and participatory. These practices empower members, promoting a sense of ownership and commitment to the community's policies and goals. By incorporating members in decisions like these, leaders like R9 and R4 not only promote transparency over decision-making but also build trust and collaboration, which is crucial for sustaining active and cooperative online communities.

6.6 Discussion

6.6.1 Human factors in community governance systems

Frequently overlooked in discussions of community governance systems are the human elements that constitute and reproduce them. No matter how one designs a framework or system for community governance, work, often volunteer-based, is required to carry out the procedures and everyday practices to achieve community goals. Prior studies on the emotional labor of volunteer content moderators on Reddit and Twitch communities found evidence of difficulty in managing toxic situations, lack of appreciation by their communities, general distrust and anger toward moderators in general, and burnout (Dosono and Semaan, 2019; Wohn, 2019). I extend this conversation about emotional labor of volunteer moderators by situating their labor more specifically in the context of *community governance* and how the messiness of human interactions and communication complicates institutional efficacy.

Emotional labor in rule enforcement

While Dosono and Semaan (2019) and Wohn (2019) highlight the personal challenges volunteer moderators face in managing their communities, my findings reveal how the emotional labor of governance in these settings impact the ways in which leaders interact with their communities and the decisions they make in rule offense situations. First, there is the issue

of inserting oneself into a situation when a rule has been broken in order to rectify it or sanction offending community members. While Dosono and Semaan (2019) and Wohn (2019) described the harassment that individual moderators faced from unhappy users who get banned or their content deleted, I find that frustrated rule violators may redirect their ire toward the entire governing body (i.e., the moderation team) of their community through public posts on their sites writing about alleged unfair treatment. Over time, leaders may begin treating rule violation situations more delicately because of fear of unwanted backlash at themselves personally or at the entire moderation team. Decisions about whether or not to intervene and reinforce community norms may become less straightforward—and rule application less consistent—as the emotions of the rule violator are taken into account. A frustrated community member may start a campaign of dissent about the legitimacy of the community’s governing body, inspiring others to lash out or pile on with more criticisms of leadership. Accompanied with harassment in direct messages targeting leaders about their governance decisions, volunteer leaders of online communities must manage their emotions as well as those of their community members to effectively govern. As these findings show, leaders took some strategies to minimize the amount of negative content they get exposed to, such as pre-emptively sanctioning potential bad actors with blocks or bans and creating alt accounts for moderating in order to distance their online identity from their roles as community moderators and admins.

Limitations of a graduated sanctions approach

Despite design recommendations and observations about graduating sanctions in rule violations by community leaders (Kraut et al., 2012; Seering et al., 2019), I find evidence of situations where this approach may not scale well with communities that grow very popular in size and activity. Volunteer moderators on a very large subreddit, for example, may find that a case-by-case approach in which rule violators of any kind are first warned before escalating to more severe sanctions is not sustainable because of the toll it takes on their own time and labor. Strategies that might resolve the problems associated with scale and minimize burnout—expansion of the leadership team and intentional turnover—

were rarely reported as in interviews except in a couple extreme situations. The resulting approach is one like R8's, where virtually any rule violation is met with the same severe sanction—removal from the community. Interestingly, this extreme approach to community governance did not also reproduce the kind of vitriol and harassment other leaders anticipated when taking any governance action against a community member at all. Perhaps the consistency in how the rules were enforced in this community was respected and understood by community members as legitimate community governance.

Interpreting situations and rule definitions isn't always clear

Finally these findings highlight the challenges that individual leaders face in identifying a rule violating situation and the definitions of written rules. Virtually most scholarship on governance in online communities takes for granted the practice of rule enforcement—that moderators and admins will always clearly interpret a rule violating situation and that the written rules don't contain ambiguities that make decision making unclear. Actually, rule enforcement may become a complex process for several reasons. First, the written rules may not account for marginal or new problems. The language of the rules may be subjective, like the widely used 'don't be a jerk' rule found in many online communities. Interpreting situations where someone's behavior may be out of line is also challenged by the lack of non-verbal communication cues (Walther, 1996) or the imagined audience context of the speaker (Boyd and Ellison, 2007a), which is detrimental to most interactions in computer-mediated communication settings. There is also pressure to get it right" in order to avoid upsetting the offending member and to minimize backlash. Incorrectly enforcing a rule may risk making future governance actions appear illegitimate and jeopardize the effectiveness of sanctions. The resulting process for rule enforcement given these factors is one that requires additional labor of volunteer leaders to gather context and information of a rule violating situation, communicate with other leaders in sense-making procedures to correctly and consistently interpret rule violation situations, and deliberate collectively on how the offending rule violator should be sanctioned. In other words, rule enforcement isn't always easy and may require more work, both cognitively and emotionally, to navigate ambiguous

situations and make decisions that optimized community cohesion.

Enforcing rules in online communities becomes less a straightforward application of pre-defined policies and more an iterative process of sensemaking. Leaders have to interpret, adjust, and sometimes retroactively constructing the meanings of rules the actions of sanctioning unwanted behavior. In ambiguous situations, leaders often rely on retrospective interpretation—drawing from past instances to guide current decision-making. As R9 and others explained explained, situations that fall into “gray areas” often require collaboration among moderation team members to determine if and how a rule should apply. These actions reinforce shared interpretations and ensure consistent or fair rule application. This practice also means that leaders not only enforce rules but actively “talk” the community’s identity into being through their interpretation of rule violation situations. By embracing sensemaking as an ongoing, social process, leaders shift from simply enforcing rules to organizing the community through “labeling and categorizing to stabilize the streaming of experience” (Weick et al., 2005, p. 410) with each decision to enforce the rules.

6.6.2 Actually existing community governance

Two perspectives have emerged in the scholarship on governance systems of online communities that are worth critiquing in light of the findings. The first perspective is that of the sociotechnical systems designer, whose central arguments are that online community can be designed and managed to achieve their goals, influenced by social psychological claims about design interventions that can hurt or hinder community success (Kraut et al., 2012). The other perspective that’s come into more recent social computing scholarship is what I may call the “idealist” perspective. This idealist perspective assumes that online community governance should strive for democratic practices but are restricted by implicit hierarchies structured into the software that hosts online communities and distributes permissions (Schneider, 2022). The idealist perspective observes ideologies of absolute power (i.e., “feudalism”) implicit in the available tools for community governance and that non-administrator and moderator users have little leverage and voice for facilitating change and alternative governing systems.

To be concise, designer and idealist perspectives take for granted the influence of the body politic (i.e., the membership base of communities) and the cost of volunteer labor in actually carrying out and sustaining community governance. While it may seem obvious to answer the question “Why *shouldn’t* members comply with the rules?”, achieving community goals is overlooked from the members’ perspectives. In other words, one should ask “Why *should* members comply with the rules?” when exit and producing one’s own community are technically viable options. Why should members buy into a framework of community governance offered by someone they don’t actually know—someone who also doesn’t really know them? Establishing one’s leadership as a legitimate power for governing an online community could require, as some of the findings suggest, the consent of the governed. Prior work has found members may hold a lack of respect or even hatred for the volunteer leaders of online communities (Wohn, 2019). Indeed, as Schneider (2022) points out, leaders of online communities are frequently self-appointed and have the capacity to rule their communities as dictatorial authoritarians with no regard for or allowance of criticisms from their members.

The designer perspective rightfully aligns with many of the arguments made by leaders we spoke to in this study that argue leaders of online communities should seek out sources to increase their legitimacy. These sources often come in the form of procedural justice mechanisms like having the chance to argue one’s case, appeal a sanction, and transparency in the governance decisions of a community (Kraut et al., 2012). However, I also find that additional work is required to legitimately govern an online community. There is of course the emotional labor involved in managing the backlash of an upset sanctioned rule violator. There is also the work that leaders take in gathering context to be empathetic to rule violators, who may just be misunderstood newcomers who can learn from their mistakes. Legitimacy isn’t just enhanced in case-by-case interactions with rule violators but has to be reinforced over time. Leaders frequently engage in transparent communication and humility about their mistakes or issues occurring in their communities. They may create procedures and spaces for members to exercise their voice in giving feedback about the state of the community, critiques about how leaders govern, and any other issue that they perceive that might be changed. In fewer cases, leaders include members in the decision-making process of

policy change (discussed in more detail in the next chapter). In sum, this paints a different picture of volunteer-based online community governance in which leadership positions don't always carry the sort of omnipotence and total control over their communities described by those from the idealist perspective.

Despite the implicit hierarchies imbued in the sociotechnical design of the spaces that host online communities, why is it that its leaders can, in some circumstances, not reflect authoritarian regimes and in other circumstances even be removed entirely and replaced with a different set of leaders by the community? I offer some explanations. First, volunteer leaders are often active participants of their communities, and some even start out as regular members with no administrative authority or moderator permissions. They may come into their position with knowledge about the experience of being a community member and what it feels like to be unfairly sanctioned or their concerns ignored. Leaders described usage of external communication channels like Discord where they formed interpersonal bonds in an informal setting with community members and leaders alike. Leaders in these situations may feel a closer connection to their community members, shortening perceived difference in their hierarchical positions. In addition to the potential for forming bonds with community members, some community leaders also designated highly visible spaces for voicing concerns about problems in the community and feedback on community governance policies. Finally, the volunteer aspect of community governance could itself impact the degree to which leaders hold onto power. There are few, if any, extrinsic rewards for being a volunteer moderator or administrator of an online community (Li et al., 2022b) and their work is often obscured and under appreciated (Gillespie, 2018; Wohn, 2019). While removing oneself as a leader of an online communities is costly in terms of having direct control over how it's managed, I find instances where pressure was put on under performing leaders to give up their positions and be replaced with leaders whom members believed would have the community's goals and interests at heart. The result of this arrangement is that volunteer leaders may feel pressured to take into account the interests of their community members because not doing so could cost them their relationships or even their positions in the community's hierarchy.

Idealist critiques of online community governance may take the lens of economist of Albert Hirschman to explain an alternative explanation for why authoritarian regimes appear

to dominate the Internet. Using the concepts of “exit” and “voice”, this line of argument suggests that users are virtually trapped within the choices of whatever communities are available because their ability to speak out (i.e., “voice”) lacks efficacy while the option to choose another community or create their own in a seemingly “open market” is usually too costly because of the challenges inherent in attracting and retaining members of a new online community (Hirschman, 1970; Schneider, 2022). Even under such environmental and social arrangements, leaders we spoke to frequently described their governance work in terms of meeting their members’ needs and facilitating the exchange of their voice in discussing problems and solutions in their communities. One could argue that online community leaders resemble the “lazy monopolist” in Hirschman’s theories about political or organizational change: exit is possible but costly and limited, which only emphasizes the importance of voice (Hirschman, 1970). However, exit is perhaps not as limited as firms, political parties, or brands (the empirical cases Hirschman frequently used) because it is actually *technically* easy to produce a new online community. Sure, the social costs are high, but exit is perhaps a more viable threat to the status quo, and managing conflicts because of a threat of exit could be overwhelming for community leaders, who are limited by the amount of time and effort they can put into volunteering for their communities. Finally, it may be the case that these actions, while opening up participation and lessening the divide of some of the power, helps reinforce the established hierarchy. It’s just enough to make members believe their values and interests are represented in governance decisions but not so much that the position of administrator or moderator loses its position and power. These types of arrangements may be defending against a regime change by offering just enough participation and voice to stave off mass member exit or dissent.

Another possible explanation for why, if exit is an option and voice can be ignored and limited, did leaders that we interviewed described a practice of championing the will of their membership can be found in Turco’s work on the communication of modern day organizational firms. One might first assume that leaders derive their authority from their position in an organization’s hierarchy and the expert knowledge they are alleged to contain (Turco, 2016). Turco takes this Weberian logic to propose that an open communication structure in an organization might undermine that authority as the knowledge of members on the

lower end of the hierarchy are considered as legitimate as those at the top. In their case study analysis, Turco found a new basis for legitimate authority was achieved with open communication in which there is mutual appreciation for each other's wisdom and trust in decision-making by leaders because they know their voices inform them. Applying the logic of the "conversational firm" to the sample of online communities in this study offers a more complex view of online community hierarchies and the role of voice in sustaining them. In fact, because volunteer community leaders are often just members who participate themselves and not exactly experts of "content moderation" and "community management" (and some explicitly admit to this in interviews), they may lack knowledge and experience that can help inform good decision making in policy changes for their communities. Therefore, they may uplift the voice of their community members and involve them in decision-making because without their knowledge, these volunteer leaders might not be able to make the best decisions for their communities.

Another explanation comes from political scientist Margaret Levi's work on how compliance to rules may be "quasi-voluntary" because while leaders can sanction, they also have to uphold an implicitly agreed upon social contract or risk dissent or revolution from their base (Levi, 1988). Voluntary compliance with rules requires members to believe that the rules and their enforcement are legitimate, and those in the designer perspective agree with Levi that participation in policy making discussions should facilitate compliance (Kraut et al., 2012). That said, leaders we spoke to described doing extra work on top of content moderation to establish trust with their members in order to increase the legitimacy of their governance actions and ensure rule compliance and member retention. They not only involved members in participating in policy discussions and sometimes had members vote on policy decisions—they also had to regularly post announcements in attempt to keep their decisions transparent and interact with members in informal settings like Discord to establish social bonds with members.

Chapter 7

COMMUNITY EVOLUTION: THE PROCESSES OF RULE CHANGE

Across all interviews, the decision to change the rules in an online community was never arbitrary or random. Leaders we spoke to described different factors that led them to change their original rule set. I break these factors down into two or three conceptual categories. The first category relates to the concept of “triggers for rule change.” In this category, leaders and community members alike observed increases in problems of participating in the community, like an increase in toxic behavior, an influx of newcomers, the proliferation of scams, and so on. Leaders also described unforeseen problems that factored into rule change, like shifts in contemporary culture and language norms, declines in member participation, and even leadership regime change. The second category relates to the concept of “rule change framing.” Under this category, I highlight the language leaders used to describe the process of rule change. We learn how leaders thought about their actions through their distinct use of verbs, like cleaning, simplifying, relaxing, updating, and overhauling the rules. Finally the third category relates to the concept of “the supply of problems.” Building off (March et al., 2000), rule change may occur as an outcome of needing to solve a problem. Several leaders we spoke to said they never had to change the rules in their community, citing a lack of problems to drive that process. I explore how the sociotechnical design of community platforms and community characteristics and conditions facilitate evolving institutional systems.

7.1 *Triggers for Rule Change*

Leaders we spoke to described (often explicitly but sometimes implicitly) about the different factors that contributed to the process of implementing a rule change in their communities. In this section, I present a novel typology of rule change factors that emerged in interviews with volunteer community leaders. This section is divided into two sets of factors that may

trigger a change in the rules. The first describes triggers that come about because of an increase in typical online community problems like toxic behavior between members, newcomers, and off-topic content. The second set of problems that may trigger rule change are potentially correlated with community age. Membership declining, leadership regimes being replaced, and stagnant initial community rules needing updating to more contemporary cultural norms are but a few of these problems.

7.1.1 Increasing occurrences of existing problems

Rules are often initially adopted to address known problems inherent to most online communities (Grimmelmann, 2017). Rule adoption however doesn't necessarily mean those problems are eradicated. Rather, those rules add legitimacy to sanctioning members who break community norms and disrupt community goals (Kraut et al., 2012). Rule enforcement helps with compliance to those norms and deter the occurrence of members breaking them. Problems may not ever fully go away, but enforcing rules can keep them at bay. That is, unless, the problems begin to increase in intensity.

Interviews with leaders of online communities revealed “triggers” (outlined in Table 7.1 for potential rule change that correspond to an increase in problems in which the initial rule set inadequately helped resolve. Leaders felt compelled to resolve an increase in these problems because not doing so may threaten the community's goals and identity. For example, when R4 took over as a moderator for their r/cybersecurity community, they said “we tried to narrow the scope of what we're trying to achieve,” by implementing rules that restricted participation not directly related to cybersecurity professionals. Through feedback from active cybersecurity professionals in their community, R4's moderation team found that quality discussions were being drowned out by newcomers asking questions about career advice or personal cybersecurity troubleshooting. Rule change was implemented to address a problem that was compromising the community's identity and making it difficult for professionals in this field to locate and discuss relevant news and issues related to their profession. FA1 had to change a rule to protect their community's goals in their Aesthetics wiki community where participants were beginning to make articles about their own “per-

sonal aesthetics.” The goal of the Aesthetics wiki is to capture information about all kinds of fashion trends, but someone’s “personal aesthetic” does not necessarily count as a trend. FA1 explained:

“I have cracked down on personal aesthetics, which means aesthetics that were made because a person wants to create an aesthetic so to speak. Because I think the aim of our wiki is to sort of record pre-existing communities and these people are just sort of making it from scratch.”

In this example, people were contributing what is essentially false information, and FA1 had to “crack down” on it because it was detracting from their goals to only record aesthetics that can be verified as pre-existing communities. A similar problem occurred in Fandom wikis where fan fiction is popular alongside the existing fan community. For instance, FA3’s leadership team had to decide whether fan art could be allowed on the Marvel wiki. FA3 said: “We had like a long, like, three or four week discussion figuring out whether we still wanted to allow fan content to mix in with the non-fan content, and ultimately we decided no.” While other Fandom wikis may have allowed fan content, the Marvel wiki decided that this type of content would detract away from the legitimacy of the information that’s shared there. These examples illustrate how leaders initiate rule changes when the community’s objectives become obscured, reinforcing them by limiting content that does not directly contribute to their achievement.

Leaders recounted instances where they amended their rules to address toxic behaviors and preserve their community as a safe space. For instance, challenges such as vandalism on wikis and circumventing profanity filters necessitated modifications to existing regulations. Furthermore, instances of targeted harassment based on identity or beliefs prompted the implementation of rules to promote inclusivity. Certain contentious topics were either prohibited or subjected to stricter moderation to mitigate conflict due to their provocative nature.

One example involved the existence of “bully users” in FA5’s community who had “been a little bit more aggressive toward certain ideologies, beliefs, identities, and such.” After calling out this behavior to these kinds of users, FA5 said those users would reply with “this

isn't written down in stone somewhere," as an excuse as to why they believed they could behave that way in the first place. FA5's leadership team immediately implemented rules against that kind of abusive behavior citing their community's goal of wanting people to "have fun", which doesn't align with attacking people's identity. From all the communities interviewed, most had some form of a rule that dealt with toxic behavior and it seemed to be part of the initial rule sets community leaders adopted.

Other conditions emerged in different communities that made members feel unsafe and required additional protections through additional rules. Safety rules varied in terms of severity. For example, leaders of communities that centered on a Fandom would try to protect newcomers from spoilers, which can ruin their experience of a fictional series. FA12's community implemented a rule against spoilers entirely and would sanction violators with a three day suspension from the community. R3's moderator team had to implement a rule against discussing content in the current or most recent season of their fandom, the show *Supernatural*, without appropriate tags to signal to newcomers that the content will spoil the show for them.

Other safety rules emerged in response to issues where community members were being scammed. For example, in R9's community, people were entering the community to essentially steal plants from unsuspecting community members. R9 explained:

"We went through a phase where there were some people when we first started getting the really big scammers, like people who are taking hundreds and sometimes 1000s of dollars worth of plants from people. We got some critiques of our rules."

After all the criticism about the lack of protections around trading plants in R9's community, R9 said, "We just finally decided that it was time to institute the stricter scamming rules." This was a rule that would require community members engaging in a trade with each other to take steps to verify the legitimacy of both parties in the transaction. The new rule stated:

"Follow the TAPLAP trade safety guide. To lower the risk of scams, we ask that all of you follow the rules in the safety guide. Most importantly, always

comment on a post before DMing, use G&S payments, check the DNT list, and always send exact photos of what you will be sending.”

Together, these examples highlight how the perceived safety of members is used as a barometer to determine if the rules and their enforcement are working effectively. If unaddressed, members may leave or participate less, which could eventually lead to the community’s collapse.

Another factor that may contribute to a search for new or revised rules is when a community experiences an influx of newcomers. Newcomers are important for managing churn and sustaining community activity, but it may also be accompanied with an increase in those who seek to do harm, troll, or vandalize members. For example, FA3 had to implement a rule to gatekeep newcomers because they experienced an influx of vandals:

“We implemented a four-day edit rule in which if you create an account, you have to wait four days before you can make first edit because we were getting so many troll accounts, so many vandal accounts, so many people sock-puppeting that it just became too much of a hassle to deal with.”

Through this strategy, FA3’s leadership hoped to make it costly for vandals by making them wait four days before they can participate in the community. The hope is that the wait might deter them from staying long enough to vandalize the community, and good faith contributors might not possibly mind to wait that long before they can start participating in the community.

Newcomers may also bring with them an ignorance about the community’s norms and history that can be disruptive, like sharing content that’s already been repeatedly discussed and generally adding noise that makes it more challenging for existing members to find quality content and discussion. For instance, due to a surge of new members excessively posting about a popular Kdrama called “Crash Landing on You”, R1’s community implemented a rule against rehashing topics that have already been covered to maintain the quality of discussions and minimize spam. R1 explained:

“So we had a huge spike in members joining. It was Crash Landing on You,

and like the subreddit was basically [just] Crash Landing on You for a couple of weeks, almost a month, where all of the posts were about Crashing Landing on You...That was a big [reason] why we implemented a longer rehashed rule. We used to be [at] 30 days, and then it got moved up to two months. I think we're at four months now."

R1's "rehash" rule had to be changed to be made more restrictive so that their community would not be flooded with discussions of old, worn out topics. R1's community is used to newcomers coming in waves as their subreddit attracts fans of Kdramas that suddenly get really popular with Western audiences. That said, they had to change the rule about reposting topics about popular Kdramas because it contributed too much noise to the community, making it challenging to find discussions about anything else.

Another problem that would sometimes necessitate a rule change was spam. Spam could be disruptive and annoying to community members. Leaders would implement rules to reduce or restrict spam entirely. Spam consisted of content like advertisements, self-promotions, welcome messages, and news articles that would basically contribute noise to the community and clutter up information. For example, FV5, who runs a Fediverse instance for musicians in the Boston, USA area, had to implement a rule against spam caused by self-promotion of artists. FV5 explained:

"I added a rule to make sure that if you're for promoting one event, just no more than two posts. I didn't want the timeline to simply be a sort of clogged artery of promoters and no people actually talking to each other. Because I wanted it to be a space for people to talk to each other."

Spam for FV5 was especially a concern because of the way that Fediverse instances display information in a single "timeline". The influx of people advertising and promoting their music event was making it harder for people to actually discuss music, and FV5 felt it was necessary to implement a rule restricting the proliferation of content that detracted from that goal.

R5 faced a different challenge with spam, this time concerning the serious topic of the

COVID-19 pandemic. R5 explained how the pandemic set off a cascade of members posting threads to their subreddit at a rate that was unsustainable:

“Everyone was just losing their shit. Posting stuff every, I don’t know, 15-20 seconds at some point. I mean, it got ridiculous. So we set a rule in place that, okay, “Mega Thread Mode” is active. So what that essentially means is that there’s one thread at the top [of the subreddit’s webpage], it’s stickied. Anything related to that subject gets pulled into that thread... With COVID, we had a lot of Mega Threads because so much information was coming through. It was just kind of, you lose quality if you’re getting someone posting a thought, you know, repeatedly. You don’t have a concentrated conversation. So that’s when the Mega Threads sort of existed in order to consolidate and kind of contain conversations and things like that.”

The introduction of the Mega Thread rule effectively curtailed spam by creating a designated space for discussions on specific topics, such as the COVID-19 pandemic. This helped funnel all related questions and discussions into a single area, preventing the subreddit’s timeline from being cluttered with repetitive spam.

Some leaders mentioned updating rules to comply with government laws like DMCA compliance (preventing the share of pirated materials) and laws related to trading illegally obtained goods. This was especially prevalent for communities that center on a fandom. R1 explained how their community had to implement a rule against sharing links to illegal sources (their No Illegal Source rule), like pirated copies of Kdrama shows. The reasoning here was that DMCA complaints could be filed against their community, forcing the moderators to take down the content and potentially get the subreddit banned from the website. R9 had to implement rules in their plant trading community about selling plants that are federally illegal under US law, like marijuana.

7.1.2 Other rule change factors

When asking leaders about when and why they change the rules in their communities, some responses stated explicitly factors like those discussed in the previous section. These factors

Category	Specific problems
Content-related issues	Increases in: - Off-topic content - Spam - Illegal content
Behavioral issues	Increases in: - Toxic behavior - Unsafe conditions for members
Membership dynamics	Influx of newcomers Member participation decline
Community maturation	Outdated / obsolete rules Shifts in broader language norms Initial problems evolved in complexity
Over-regulation	Overly complex/exhaustive rule framework Stifled participation
Leadership and administrative changes	Moderator turnover / regime change

Table 7.1: A list of problems that may trigger a change in the written rules of an online community. Some of these problems may only trigger a change when their presence increases to the point where it significantly interferes with community goals and member experiences. Other problems are triggered as leaders audit their governance practices and documents and evaluate their efficacy and how well they address modern community problems.

were easy to identify and explain. On the other hand, some leaders described the process for rule change without pointing to specific instances of problems in their communities but changed the rules anyway. I derived additional triggers for rule change by analyzing the language of rule change in these responses. Leaders choice of specific verbs in describing rule change—such as “cleaning”, “fixing”, “updating”, “relaxing”, and “overhauling”—revealed underlying problems not directly or specifically caused by the content or behavior of members in their communities. Instead, the problems implied emerged over time as the initial community rules aged in ways that made them inadequate for modern community problems. In other words, rule change may not just be triggered by specific instances of a particular problem but also when leaders stop to evaluate whether their initial rule set aligns with current community needs.

Some leaders referred to the process of changing rules in a way that resembled “cleaning” or “fixing” problems with rule changes. FA3 remarked heaving “greatly tried to fix all the kinds of problems” that existed since the beginning of their community’s creation but had not been addressed. At one point, a moderator of FA1’s community suggested they “should fix the rules and clean up a lot of the pages,” which followed with meetings to discuss how to do so. FV11 changed rules “to make them clear,” and R7 saw their rules get rewritten. In these examples, the problem that initiated changes to the rules was centered on how the rules were communicated and interpreted by their perceived audience. Members of their community had not complained about a lack of clarity in the rules, and for a time, these leaders found the written rule language adequate for describing the norms in their communities. Nevertheless, they initiated rule change after some time because they felt the initial writing could be “cleaned up”, “made clear”, or “fixed” because of perceived deficiencies. These are perhaps not consequential changes to the rule sets as the rules themselves largely reinforce the same norms and address the same problems.

On the other hand, other leaders referred to a process of rule change as “updating” or “adapting” the written rules in some way. FA5 said they’ve been working on “updating” their rules because “some of them are a little outdated.” FV2 tracked rule changes on Mastodon.social and would update their own instance’s rules with whatever updates occurred on Mastodon.social. FV2 explained “the no illegal content in Germany is new on

their rule set. That wasn't there when I created my rule set, so I guess I have to now change my server rules." R3 described an instance of rule change saying:

"The rules have changed where we adapt to the changing environment of the show. When the show ended, we changed some of the rules when the show was still airing, we had to update the rules based on stuff happening."

In this example, R3's community had to adapt to changes happening with the media on which their fandom is based, requiring an update to the rules. FA3's community instituted a rule to not "dead name" a person who transitioned into a new gender identity, which required retroactively and proactively removing a transgender person's name assigned at birth from any articles on their wiki. FV16 remarked how one of their rules on harassment evolved over time, describing how it became "a little more detailed and specific over time, as we've experienced more specific forms of it to deal with." New kind of harassment emerged that weren't previously considered and had to be added to the language of their harassment rule to deter it from happening. In general, rule change approached this way implied that previously solved problems had evolved in complexity. Rules reflected a different point in the community's history that didn't match with their current needs and how problems, too, had aged in unforeseen ways.

Another way leaders framed rule change was as a process of "relaxing" the strictness of already written rules. Relaxing may either have meant reducing the amount a rule is enforced or getting rid of a rule entirely that was restricting certain forms of participation. For example, FA5's community decided to relax their enforcement of behavioral rules saying their community was "feeling kind of suffocated" and that they decided they maybe "need to not be so hardline on handling these things." FV5 remarked on their philosophy of why rules may be added, saying they "might have to take rules away if they seem overly restrictive." This happened for R3's community when the show their community is based on had a season finale and activity started to diminish on their subreddit. R3 explained their approach to the problem:

"We had a discussion about—okay, are we going to relax the rules? Like how

are we going to really? How far we going to relax them? Are we going to allow people to post fanfiction? Or even allow people to post memes?”

R3’s moderation team debated which rules they were going to remove or stop enforcing in order to encourage more activity and participation in their community.

Implied here are two problems that might trigger rule change. The first is the problem of excessive regulation. Rules may have initially been adopted to structure participation in a particular way, but too many rules restricting how a member can participate may actually deter people from participating at all. The second problem implied here is that of community decline. Online communities experience waves of activity and participation, often referred to as a “rise and decline” (Halfaker et al., 2013; TeBlunthuis et al., 2018). While the community was not necessarily excessively regulating content, the removal of the “no memes” rule encouraged members to participate in ways that weren’t relying on new content from their fandom. Removing a rule that was restricting a kind of content from being contributed during periods of decline may help a community survive.

Finally, the last framing of rule change emerged through the use of the term “overhauling”. Leaders described instances in their communities’ lifecycles where they put in effort, often in collaboration with other leaders of their community, in completely rewriting all of their rules. FA4 said “it was to overhaul our own guidelines to properly consolidate all the unwritten rules that we have,” and to “ensure not only the proper behavior from contributors but also that their contributions are up to standard.” In other words, FA4’s team decided to formalize norms that emerged over time in their community as written rules and also to improve the contributions to their wiki and establish standards. In FA12’s community, the rules were overhauled “because it was just like really complicated and kind of difficult to keep track of [the rules].” Leaders would often revise rules to make them clearer or more simple and make their job in enforcing rules easier. A similar example was stated by R9 explaining an upcoming change to their community’s rules: “We’re about to completely overhaul the review system, which is going to make it really better.” This new requirement for members to leave detailed reviews rather than just star ratings followed complaints from community members about one-star reviews with no context. R9 believed that this change

would decrease the need for their moderation team to intervene, as community members would essentially be moderating each other.

Some problems that would trigger rule change were implied in leaders' responses about rule change "overhauls." First, leaders may not fully understand *what* the norms actually are for their communities, or norms may emerge after the fact of community and initial rule creation. By not formalizing these norms into written rules, the community may not succeed at producing high quality information. The other problem implied in these responses is that the existing rule set may be counter-productive to community governance. In other words, an exhaustive or complicated rule set may take up too much time and energy to adequately enforce. By overhauling the rule set with a more simple framework, leaders could more efficiently govern. This was especially important as the act of online community governance in these settings is entirely voluntary and typically performed by a handful of moderators and administrators.

7.2 Supply of Problems

So far, I have categorized the various ways leaders described factors and approaches to rule change in their online communities. Another pattern that occurred for some leaders was an almost total absence of rule change. For some leaders, the absence of rule change reflected an absence of serious problems in their community. FA6 explained how in the early stages of their community's life, the community: "didn't have rules at the beginning because the number of edits was in the very low volume. So the number of times that people would break the rules, either intentionally or mistakenly, it was really low." The need to create rules did not even register as a necessity until their community grew larger and started seeing more activity, including people that were contributing low quality content or vandalizing. A community's population size and age may correlate with a supply of problems too low to initiate a rule change. Communities in the early stages of development may be not only small but homophilous. In FV17's community, they explained how in the initial stage of their community, "there weren't that many problems because we kind of know each other already" and that they "didn't have to change any of the rule" because they were not having to actively govern in their community. In other words, the fact that they "kind of know each

other already” means that members are similar in their beliefs and approaches to interacting online and don’t perceive problems as being requiring a change in rules. Finally, the size and makeup of a community could be controlled through gatekeeping infrastructure, like the open or closed registration options available to Fediverse instances. FV10 explained that after adopting their initial rule set in the beginning, they “didn’t do a whole lot more from there,” because their Fediverse instance “pretty much never had open registration.”

For others, leaders may have been constrained in their capacity to know what problems exist at all and to address them. For example, FA6 explained why there wasn’t much to change about their rules “because at the time, we were more preoccupied with fixing the wiki.” Leaders of wiki communities have extra responsibilities beyond moderating content—they may have to clean up vandalized pages, make sure that the pages and table of contents are organized appropriately, or are busy creating articles themselves. Such responsibilities may take attention away from problems with the rules. Time is a valuable resource for volunteer leaders of online communities, and when it’s low in supply, leaders will prioritize immediate responsibilities rather than meta projects like rule change. R1 explained why the previous leaders of their community could not get around to evaluating and improving the rules because “the other previous mods never really had the time,” and were “kind of busy with real life.” The voluntary nature of leadership in online communities means that many leaders are only active when they’re not working or dealing with their personal or family life. Responding to reports may be the extent of their capacity to govern. That said, R1 also explained how the COVID-19 pandemic, which required many workers to work from their homes as well as restricted outdoor and public activities, granted additional time to their leaders, expanding their capacity to govern. R1 explained:

“There was a huge amount of time where all the things—all the time I used to devote to other real life activities, like things that are not in front of a computer, all of a sudden, I had so much free time, and I also think for all of the mod team. So, all of a sudden, we had that free time to really work on the community, and that’s helped.”

It was during that time that R1’s community finally worked on overhauling their rule set

and figuring out how to encourage community interactions. This example highlights how the voluntary nature of leading online communities constrain their ability to address ongoing problems and improve the community's experience for its members. However, even when leaders have time to address a problem, they may be limited in their capacity to enforce a new rule. FA1 explained how they decided not to change a rule on their wiki because it would require them to retroactively edit article pages so that it would align with the new rule, FA1 said, "if it was like a large mass deletion or mass editing that involves a lot of writing, then that would be a lot more difficult to coordinate." Rules aren't instant solutions to problems, and their addition may create more problems for those who have to enforce them. Online community leaders face an interesting challenge as they are both the enforcers and creators of these rules, and these dual responsibilities may constrain the evolution of their community's governing arrangements.

Overall, rule change events seem to correspond with the way that leaders view a supply of problems in their communities and their capacity to effectively resolve them. Problems are essentially socially constructed. As March et al. (2000) explains, "organizations are thermostatic in their attention: as long as performance is satisfactory, problems tend to go unnoticed." (p. 63) However, these findings diverge from March's proposition that demand for problems increases the potential for pattern recognition rather than the supply. In other words, March believed that a low supply of problems would make minor annoyances seem more severe, initiating a change in rules to resolve those problems. Unlike policy writers for profit-seeking organizations that March studied, volunteer leaders who both create and enforce rules are constrained in their capacity to recognize problems as well as work toward resolving them.

7.3 Rule Change Process

The emergence of problems doesn't always result in a rule change so long as the organization is operating at a satisfactory level (March et al., 2000). I found that online communities are organized in a similar way when it comes to the process of rule change. The existence of a problem might result in a rule change after certain conditions have been met. First, it has to gain the attention of community leaders. Attention to a problem could be first hand as

leaders observe a problem in their interactions with the community. Attention might also be achieved through members raising their voice about the problems they experience in the community. Once leaders are made aware of a potential problem in their community, they may begin a process of resolving it through a rule change. This process may involve a series of deliberative discussions about what the problem actually is, which change in the rules could address the problem, and how to formally write the new or revised rule. Leaders may seek out additional context and ideas from their constituent community members and iterate on resolutions collaboratively between their leadership team and the community. Finally, a decision has to be made about whether to implement the rule change, and the ways in which these decisions are made may depend on the governance models of the community (autocracy, hierarchical oligarchy, flat oligarchy, or limited democracy).

7.3.1 Voicing concerns to leaders

Leaders described a process for facilitating rule change that generally involves first someone expressing their concerns about a problem in the community. This then reaches the leadership who deliberate about how to resolve the issue, if a resolution is needed, and they frequently get community input on the matter before making a final decision. Leaders assume that members will feel comfortable enough to voice concerns to them and that member problems are valid. That said, space has to be made to receive member complaints and feedback as well as to get widespread confirmation that a problem needs a rule change. For FA1's community, members were encouraged to directly message their leaders under the logic that "if there's a way to fix it, then you can always discuss it with the moderators and then they'd agree to a change." In FA1's case, problems could easily be addressed so long as members took effort to point them out to the leaders. Responding to alerts about problems with a rule change was relatively straightforward for communities like FA1 and FA11. FA11 said "rules are usually added whenever the user base complains or my admin team complains or my discussions moderator team complains." FA11's description of the process implies that they view feedback as "complaints" and that the addition of rules in their community usually resolves whatever is being complained about.

Some leaders described slightly more formal mechanisms for feedback about problems in the community. For example, R4 noticed that members of their community were using the report feature on Reddit to signal problems about a kind of behavior in the community to the leaders as well as creating threads to voice concerns about it to other community members. FA3's community also utilized forum style posting for creating spaces for feedback about problems in the community. FA3 explained that "if someone has an issue with a rule, they bring it up—the forum, the discussions—to an admin." R1 also stated that they changed some rules as a result of getting a lot of feedback from community members in discussion threads on their subreddit about a problem related to types of low quality and off topic content people were sharing.

7.3.2 Deliberating on how to change rules

Leaders, especially those in teams, often come together to discuss problems observed or mentioned in the community to come up with solutions that might require a change in governance (i.e., a change in the rules). The process may involve first becoming aware of a problem, bringing the problem up to other leaders in a private communication channel, individually contributing ideas for solutions, deliberating about the potential solutions, and then finally deciding on what should be done. These discussions are not always perfect or productive, with some lasting hours and may even require phone calls rather than synchronous text chat. Some leaders draft and discuss potential rule changes, while others may take a potential solution and test it for edge cases and gray areas that might make enforcing the new rule challenging. Teams had to make sure these discussions were safe for all participants so that they could communicate disagreements and avoid a dictatorial situation where one person dominates the decision making.

Despite their volunteer status, leaders we spoke to reported having to spend much of their time in meetings with other leaders about how to resolve problems in their communities with rule changes. FA1 said their team of moderators had three hour meetings to try to "fix" and "clean up" their rule pages. FA3's team once had to deliberate over three to four week discussions as they tried to figure out if their Fandom wiki would allow fan created

content to mix with canonical content related to their fandom. During a guidelines overhaul in 2019, FA4 was involved in “a very lengthy process with the rest of the administrative team” in which they they worked on drafting an updated version of what they considered proper behavior and contributions to their Fandom wiki. R9’s team once had such a serious issue with a scammer in their community that a phone call between all the leaders was required to deliberate how to resolve that issue. Overall, leaders considered these meetings and discussions with other moderators and administrators as part of the process for figuring out how to solve a problem with a rule change.

Deliberation within leadership teams often allowed for different points of views and space for ideating solutions to the problems in the communities. R9 described a process for deliberating with other leaders that reflected a respectful environment for communicating ideas:

“Like, we want to make sure we all get along well, so that we can communicate disagreements and not have it be like a big thing in the group. And usually, we’re able to get people to kind of mostly agree or be like, yeah, I have a different opinion, but I don’t care enough, so like, go for it, and stuff like that. But if it’s something that two people really care strongly about, usually they’ll just end up talking to each other enough. And they’re respectful enough that they’ll get to the point where one of them is swayed one way or the other. Or it’s just tabled for another month or two.”

In R9’s example, leaders can disagree but their norms are such that it doesn’t disrupt the group because people are respectful. If a consensus can’t be met, they simply revisit the issue later to allow more time to pass. FA6 described their process for rule change deliberation and how it involved testing for “edge cases” to make sure that the rule change was informed with as much context and nuance as possible:

“So the way that we make sure that’s fair is that, once...we’re making the rules, it’s when a lot of that deliberation comes in...When we propose a rule, or we’re like thinking through it, I tried to test it for like, edge cases, like the gray areas

to see how we do that, or whether we have to change the rules or whether it fundamentally makes it like not very enforceable.”

FA6 gave an example of this with a potential rule change about banning “leaked content” related to their Fandom’s media, the video game Valorant. During a meeting about the rule change, FA6 showed examples of what could be considered leaked content but is actually official information from the Valorant company, just published in a different region. This context was then included in the write up of the formal rule and how it would be enforced.

Other leaders we spoke to also described problem solving with rule change as a collaborative effort. R7 said “it’s more or less in a group chat—kind of everybody adding their own cents to it and then changing [rules] as a group.” R5’s community was rocked by Alabama’s abortion rights laws changing, which prompted their team to get together to try to figure it out. R5 said that “when a topic gets extremely popular, we sort of got together and sort of figured out how do we handle when a topic has gotten hot.” Many leadership teams utilize other apps like Discord to facilitate these leadership discussions. FA11 said:

“Most of the discussions are happening in the Discord server that we have and specifically the wiki administration channel. So usually there’s a discussion had there about changing the rules, and then usually I draft it up and then we’ll discuss it.”

FA11 described an iterative process where information is gathered in a discussion between leaders, FA11 then drafts up a proposed written rule change, and then that goes back to the leadership team to further discuss and iterate on the language of the rule as needed.

7.3.3 Including community members

Although leaders may then decide to implement the rule change, in many cases, leaders we spoke to described circumstances where they sought out additional information and context from their community members before making a final decision. Some even had community members weigh in on the problem and if they would be okay with a rule change to address it. For example, when planning to add additional rules to their community, FA6 said “we

actually did outreach for a long time,” noting their efforts to see what community members wanted and how they would react to a rule change.

Many leaders emphasized the importance of including community members in these deliberative discussions and valuing their input. FA3 remarked on their process and how it prioritized community feedback:

“We try to get great input. We try to get input from other users. And if we come to a decision, and everyone else is like, ‘No, that decision sucks,’ we’re like, “Well, we think it works better because this, this, and this, but we see your point, and we can compromise here, here, and here on that change.”’

Here, FA3 calls the feedback “great input” and demonstrates their leadership team’s capacity to compromise when community members disagree about the way that a rule change would be implemented. When initially setting up rules in their community, FV17 used a shared Google Doc with a draft of the rules so that “anyone who wants to—who saw that link—could give suggestions.”

Community member input was gathered in rule change discussions in different ways as well. R4 described a structured process for gathering community input about the problems in their community, the rules, and potential rule changes:

“We just built on [the rules] iteratively, with consultation with the community saying, you know, this is a problem. We say, we’ve seen people flagging these other posts, or people complaining about this type of behavior. This is how we intend to deal with it, with the change to this rule and that rule.”

R4 seeking out “consultation with the community”, along with the fact that their team does this process almost every month, implies that rule change is an iterative process as well that builds off both member and leader input. Leaders like R5 utilized their subreddit’s discussion format to get a pulse on how the community felt about potential rule changes to discussing controversial topics like the Alabama abortion ban. R5 said “when we make major rule revisions, when the mod team got together, we talked to the community and got their feedback.” R5 described the process:

“We just did a thread, open discussion...I don’t think we did a poll or anything like that, it was more open discussion than looking at a vote. To me, we sort of use upvotes as this is something you’re supporting—our opinion—you know, react or whatever. So we sort to just let it filter itself out naturally and then looked at the quality of discussion for what people were wanting.”

Here, R5 saw how the community members utilized Reddit’s “upvote” and “downvote” system to show support for opinions related to the potential rule changes. More upvoted comments in a discussion get a score and rise to the top of the discussion thread while downvoted comments get a lower score and fall to the bottom of discussion threads. In this way, R5’s team was able to gain some understanding about how the majority of the community felt regarding the proposed changes.

One community leader, R8, had a dissenting opinion on the use of gathering feedback from community members, especially through formal processes like surveys. R8 said:

“We have a Google form that we have up to get feedback from everyone...but every time that I’ve done that, I’ve never really gotten any actual concrete feedback on serious problems.”

R8’s experiences may reflect a challenge in gathering input from community members, or it could also be the case that members who responded to the survey were mostly quite satisfied with how things were operating. Nevertheless, this example highlights the possibility that opening up participation in governance changes to members isn’t always a useful endeavor, and because these leaders are volunteering their time to manage these spaces, it may even be an inefficient strategy for figuring out how to improve community conditions.

7.4 Decision-Making Structures

Once information has been collected, analyzed, and potential changes drafted, a decision must be made whether or not to implement the rule change. Through interviews I discovered that different leadership teams had different structures in place for making these kinds of governmental decisions. Some centralized decision-making into a single leader, others would

find consensus within their small group of leaders, and even fewer would put the decision up to a vote of the entire community collective. While each sociotechnical setting I investigated had some variation of these decision-making structures, it became apparent through the analysis of interviews that each one seemed to lean more toward one particular governance model than the others.

7.4.1 Autocratic decision-making in the Fediverse

Over all the different sociotechnical settings from which we interviewed leaders about rule making, it became apparent that Fediverse instance administrators were more often than not the sole leader of their community and consequently the sole decision maker. Most of the time this was implied when Fediverse leaders replied in the first person, such as FV3 answering a question about whether their community's have ever changed, saying "But the rules themselves? I haven't touched them in a long time," or when FV11 said "I think I've added—maybe...tried to make them clear." In one interview, FV1 said, "the rules as such reflect what I want personally in a community experience or social media experience." These interview quotes are meant to illuminate that when talking about how leaders on the Fediverse govern their communities, they tend to view it solely from their perspective and as their sole responsibility. This perspective aligns with much of the ideological underpinnings that make creating communities on the Fediverse appealing: technolibertarianism and autonomy over one's social network experiences. The resulting decision making structures for Fediverse instances then begins to resemble autocracies. A single leader, typically the administrator who created and pays for the hosting of their Fediverse instance, is the sole decision maker about governance over their community.

Coincidentally, I also saw little about changes in rules in Fediverse instances. Most "problems" reported by Fediverse leaders we spoke to were related to external users and other instances with which their community was federated. A change in rules only affects the behaviors of members within their own Fediverse instance and would not have any effect on the way people in other instances behaved. In other words, rule change seemed rare because internal problems were virtually nonexistent and because problems, typically

resulting from interactions with other Fediverse instances, were resolved through the action of “defederation”, which is when a Fediverse instance closes off their network to the offending other Fediverse instance.

Finally, growth did not seem to be a priority for the Fediverse leaders we spoke to. Indeed, the motivating factor for the Fediverse was the belief that you should have ownership over your social network experience, and this encouraged people to create their own instance if they didn’t align with the norms or rules of any existing communities on the Fediverse. Leadership teams as a result remained relatively small, if not entirely solitary with a single administrator finding it sufficient to do all the governance work for their Fediverse instance. All together, these factors, a low or nonexistent supply of internal problems, controlled or nonexistent community growth, and unitary leadership, resulted in few changes in community rules. When rules did change, they were without much deliberation and often a result of simply updating their rules to appear legitimate to other Fediverse instance administrators.

7.4.2 Oligarchical decision-making in Fandom wikis

Most leaders from Fandom wikis we spoke to described a decision-making structure that was consolidated to a small group of leaders, like an oligarchy, that manage their respective community spaces. This became apparent changes as many Fandom wiki leaders used the term “we” in response to questions about how they create and change their rules, referring to their team of leaders. When it comes to making rules for FA3’s community, “ultimately, the decision is down to the admins and content moderators.” Even after they get input from community members, the decision to make any change in their governance is left to a small group of leaders. That said, internally, these decisions may be made collectively among the leaders or from a central figure. For example, FA3 said, “we’re a democracy, so it’s not like a dictatorship where one person decides things, but we have a small collective of moderators and admins that make decisions.” In communities like FA3, leaders deliberate and seek consensus about how to resolve problems with rule change. FA14 said, “you always try to compromise or you try to make your interests—make your points to have these rules [or] not have these rules.” In other words, it was understood that while all leaders’ voices

were considered in discussions about rule changes, a consensus among the collective was necessary to enact any changes.

Other Fandom communities had a more hierarchical internal decision-making structure. FA1 said that the original creator of their community changed their rules without deliberation with the other leaders: “she was on hiatus for half a year before she returned, and decided that the rules were too lax.” Despite being away from the community while FA1 and others managed the wiki, the original leader had the authority to make drastic changes to their community’s rule set immediately upon return. In FA11’s community, a more formal executive role emerged in which single member of the leadership team was selected as the “main bureaucrat” who was given the authority as the “final decision maker.” FA11 described the main bureaucrat as having “the central role of helping make the rules for the wiki and making the final kind of decisions if there was like a tie or anything like that.” The main bureaucrat had the exclusive authority to draft rules for the community and could also use their position to resolve ties when the leadership team was voting on how to change rules to address problems.

7.4.3 Limited democracy in subreddits

While most subreddit leaders we spoke to described similar oligarchical decision-making structures to those discussed above, a couple leaders (R4 and R9) described a more democratic approach to the implementation of a rule change. For R4, this was embedded in their philosophy to managing online communities. R4 said, “I think the good way to handle rule development is based off the community’s needs and consult them the whole way.” R4’s leadership team would frequently send out surveys to their community members, create meta discussion threads inviting members to give feedback about the state of the community and its rules, and would take that feedback to change or improve the existing rules in light of the community members’ stated needs. In this way, the leaders of R4’s community see themselves as facilitating the will of their community members and do the work to craft and change the rules based on what the members need the most.

For R9’s community, they changed rules based on a similar philosophy as R4’s: “We try

to change things as the community wants [things] to be changed.” Before R9 joined their community’s moderation team, the community was dealing with a team of moderators that weren’t taking actions in the best interest of the members. Those leaders were forced out and replaced with a new team of leaders, including R9. As a result, leaders of R9’s community try to be “as hands off as we can while also keeping people safe.” This means that when it came to changing the rules that could disrupt or upset community members, R9’s leaders are cautious and said they had to “wait until we absolutely had to institute a rule like that since so much of the community didn’t want it.” Decisions that could disrupt the community require deliberation, input from community members, and potentially require member consensus on a vote about whether or not to implement the proposed rule change. R9 said:

“We do community posts. We try to do them once a month. We have polls. So if we’ve seen people a little bit cranky about some rules that we have, we will put out a poll...to see if they want things to change or not.”

When a consensus can’t be reached, such as a “50/50” tie, the proposed rule change is doesn’t succeed and the rules remain the same. All that said, voting to change a rule is only sometimes carried out in R9’s community, and at the time of the interview, R9 said that a major rule overhaul was being planned to simplify moderation responsibilities.

7.5 Discussion

In this chapter, I described the complex processes of rule change within online communities. First, I focused on the factors that initiate change and how leaders perceived both the problems in their communities and how they framed the process of resolving them with a change in the rules. I then described the process that goes into formally changing the written rules in an online community. Leaders made sense of the problems they were facing, may seek out additional information from the community, deliberate with other leaders or even with community members, and finally make a decision on whether or not to implement the rule change. I mapped out the entire rule change process in Appendix G.

Analyzing the findings produced insights that could help explain the variation in outcomes of the different communities’ histories of rule change in this study. First, I examine explanations that may help explain why rules *did not* change in some of the communities in the data. I then pull on two different theoretical strings, one based on March’s “supply of problems” dilemma and the other based on Hirschman’s “exit” and “voice” concepts, to situate the variation in rule change rates. I discuss alternative factors for rule change that emerge when communities face existential pressures to change from being in periods of decline to explain why they may completely evolve their institutions. I follow then follow with considerations about the sociotechnical design differences that may have also contributed to or occasioned different outcomes.

7.5.1 Community problems don’t always trigger a rule change

Although seemingly obvious, the fact that the existence of typical community problems do not always trigger a change in the rules is worth considering in light of recent scholarship on factors that may contribute to a change in rules or digital institutions (Frey and Sumner, 2019; Reddy and Chandrasekharan, 2023). These factors, age and community size, are grounded in past work on rule change in formal written institutions (March et al., 2000). An organization like an online community, over time, may come across new problems that can be solved with a policy change, and when it grows larger, problems may become intensified and require more complex rules (Frey and Sumner, 2019). Reddy and Chandrasekharan (2023) found statistical evidence of rule change on Reddit that showed community growth being positively associated with rule change while community age being negatively associated with rule change. Their findings about age contradict Frey and Sumner (2019), but they might be explained by the issue of a rule regime (i.e., those in power of creating rules) aging may not see problems as starkly as a brand new rule regime.

A possible explanation for the lack of rule change in the Fediverse sample in particular has to do with the function that rules serve for Fediverse instances. In their work on rules and the evolution of governance on Wikipedia, Butler et al. (2008) found that rules can serve a wide variety of functions, including the function of constructing meaning and identity (i.e.,

“who we are” and “how things should be”) and the function of being external signals to outside stakeholders (i.e, symbolic of legitimate governance). In other words, rule may just serve as building a “buffer between themselves and [the organization’s] environment” (March et al., 2000, p. 60) to show symbolic compliance to external stakeholders (Meyer and Rowan, 1977). Similarly, I discussed in an earlier chapter how rules are often adopted by Fediverse instance leaders to signal legitimacy to potential newcomers (advertising what the community is like, what it stands for, etc.) and to signal legitimacy of content moderation to other Fediverse leaders in order to avoid defederation. Recent work on Fediverse rules from Nicholson et al. (2023) also found that Fediverse rules predominantly focused on behavioral rules rather than rules constraining things like topics, formatting, and other content related rules that are normally found on rules on Reddit or wikis. Problems for Fediverse leaders are usually solved in either of two ways. Internal problems from members may just be kicked from the Fediverse instance, and external problems from users of other Fediverse instances may just result in individual blocking of the user or defederation entirely from a perceived “bad” instance. Furthermore, Fediverse instance admins may just have a lower demand for problems (discussed in the next section) compared to leaders on another platforms like Reddit or Fandom.

7.5.2 Institutional sources of change

Theories of institutional change can help explain variation in whether leaders may decide a change in the rules is needed. I draw on two different lines of work. The first takes the issue of rule change as a result in leaders not just noticing problems but that there is a demand for the problem to be resolved with a change in the rules. The second examines the way institutional arrangements and environment constrain or facilitate collective action in demanding change from leadership.

The supply and demand of community problems

March et al.’s (2000) work on the evolution of written rules considers how “organizations are thermostatic in their attention” to explain how leaders of organizations make sense of

problems and whether they should be solved. In other words, as long as the performance of the organization is satisfactory, problems, while probably existing, tend to go unnoticed by leaders (March et al., 2000). Problems are therefore “socially constructed” in the sense that they become serious matters only when there is a demand to notice problems at all. Online communities may have problems that don’t have to be addressed until performance of the community (i.e., maintaining high quality and active participation) begins to suffer. R9’s recount of delaying the adoption of a widely used “goods and services verification” rule is a good example of this. It wasn’t until their community experienced growth in membership that the problem of managing reports about fraudulent trades became too make the verification problem more acute, requiring the adoption of the rule. However, the fact it wasn’t put into place anyway is worth considering. The demand for changing the rules might be lower when the cost of changing them is too high (Ostrom, 1990; North, 1990). In this case, a verification rule would require additional work for members to trade with each other, which could deter participation in their community.

Another overlooked source of demand for the attention to problems comes from the fact that leaders we spoke to, unlike those in formal organizations, are volunteers who are not only charged with creating rules but also enforcing them, which takes up a lot of their time. Volunteer content moderators put in considerable work in their communities, both in regulating content, communicating to members, organizing the moderation team, training new moderators, creating policies, and participating in the community (Grimmelmann, 2017; Li et al., 2022b; Dosono and Semaan, 2019). When small groups of volunteers are already burdened with moderating content in their communities, there may be even less time available or become burned out entirely from how much work they have to put in to their communities (Schöpke-Gonzalez et al., 2022). The resulting situation is one in which volunteer leaders of online communities may experience times where there is less demand for problems as the everyday practice of governance is already taking up much of their attention. For example, R1’s community didn’t re-evaluate the state of their rules until the COVID-19 pandemic forced many of their volunteer moderators to work from home or become unemployed, which coincidentally allowed them to have more time to evaluate the state of their community’s rules and consider changing or updating them.

Effective exit and voice mechanisms

Another body of work exploring factors for institutional change comes from Hirschman's (1970) concepts of "exit" and "voice." This perspective examines the choices that consumers (or organizational members, citizens, etc.) may take to signal deterioration in the quality of their respective product, organization, state, or whatever group or audience in which they're apart. Signaling deterioration when one is not in control of decision-making often means one of two strategies: "voice", in that one may communicate their displeasure or criticisms with the status quo while remaining part of the group in hopes that it will result in change, and "exit", which is a non-verbal signal where one leaves the group as symbolic protest of its deteriorating conditions. Members can be limited in their capacity to exercise either of these strategies given the institutional arrangements in which the group is embedded. For example, if a group has a monopoly on a certain audience, exit may not seem limited if there are no viable alternative groups to join. Likewise, change may be less likely if no formal mechanisms or spaces are developed for facilitating effective voice (Ostrom, 1990; Frey and Schneider, 2021). This perspective is useful for understanding why change may occur in some online communities while less frequently or never occurring in others.

I find that, like Frey and Schneider (2021) suggest, that all three sociotechnical systems in this study, Reddit, Fandom / Wikia, and Fediverse platforms, are not designed for facilitating meaningful mechanisms for exercising voice to affect institutional change in communities. Embedded in the design of each are essentially oligarchical or autocratic decision making systems where leadership cannot be systematically challenged and replaced. In addition to this point is that the low cost of producing *new communities* on each of these platforms means there is a possibility and in many cases a reality that other communities exist as options for exit. This arrangement is what Hirschman may refer to as a "lazy monopoly" in which leaders may not have to change policies when members are upset with the status quo because those same troublesome critics are welcome to leave and join other communities. Even when members point out that problems exist, leaders of online communities like the ones I studied are not obligated to resolve them.

I do find, however, that these sociotechnical arrangements differ in the possibility for

members to organize against and change a rule regime. After all, I described policy change and how it occurred in different communities in just the sample from this study, some changes even resulting voice mechanisms in which community members expressed criticisms with the status quo and leaders followed through with appropriate rule changes. In other words, it isn't necessarily the case that leaders are bound to whatever implicit ideology structures hierarchies in online communities. Instead, it might be more useful to understand how and why communities push out from status quo oligarchical control and open up to democratic participation and meaningful voice mechanisms and evolve.

7.5.3 Evolutionary dilemmas as occasions for rule change

I offer an alternative explanation for rule change that takes into account existential pressures for online communities to change their policies. Drawing on Iriberry and Leroy's (2009) community life cycle perspective, the history of an online community goes from inception to creation to growth and either matures or declines toward death. Each stage of a community's life cycle may require a different strategy and technology to see the community thrive, but this often depends on either the designers of the systems that host online communities or the discretion of the leaders who manage them. When participation decreases and quality begins to go down, communities that don't alter their strategy may decline toward death (Iriberry and Leroy, 2009).

While recent scholarship focuses on factors like community age and finding positive associations with community size with rule change (Frey and Sumner, 2019; Reddy and Chandrasekharan, 2023), interviews reveal an under-theorized problem related to community evolution: that of community decline. While policy changes to improve quality might also lead to less participation and decline in communities (Halfaker et al., 2013), examples from the interviews also show that in some cases, doing nothing when a community is deteriorating could jeopardize its future as members lose faith in leadership to effectively govern and manage problems that are harming community goals. Communities in decline may be able to rectify their situation with incremental changes, like when R3's community noticed less participation during a period of no news for their fandom and removed a rule that

prohibited the sharing of (albeit low quality) content of memes in order to spur growth in activity and retain members. Incremental rule change is often less costly than completely overhauling the rules (Ostrom, 1990), and it would seem that communities that rely on members to thrive (wikis and subreddits in particular) would regularly evaluate the state of their rules and how well it's working in retaining and attracting members. That said, the findings show that replacing an entire institution (i.e., a rule set) may be worth the cost when problems fester for too long, leaders fail to acknowledge or resolve them, and members collectively voice dissent and demand action. In cases where rules are completely overhauled, they may also be accompanied with a regime change—turnover of the old leadership team and replaced with a new one. New leaders may be more alert to problems (March et al., 2000) and willing to experiment with new rules given the status quo rule set was insufficient in helping the community thrive beforehand. When rules change dramatically, like the overhaul described by R9, it may be influenced by pressures to save the community from permanent decline and eventual death.

7.5.4 Technical opportunities and constraints for rule change

So far, I have discussed the social and institutional arrangements that limit communities in their capacity to change policies and evolve. However, I also observed from interviews that recounts of change may occur more for Reddit and Fandom communities than for those on the Fediverse. The differences in the sociotechnical design of the platforms that host communities may have helped facilitate higher rates of change in some and lower rates in others. Past scholarship has theorized and examined the ways in which technology design will provide opportunities for or constraints on actions of those who use it (Leonardi and Barley, 2008; Kraut et al., 2012; Orlikowski, 2016). A materialist approach described by Leonardi and Barley (2008) would help us understand how the technological infrastructure in which these communities are embedded resist the kind of change that would allow communities to evolve. Each setting where we sampled communities represents a radically different technology experiencing similar problems of governance but reaching different outcomes.

Scholars have insisted on examining the ways in which technology design provide oppor-

tunities for or constraints on actions (Leonardi and Barley, 2008; Schneider, 2022). Specifically, how designs can give rise to more democratic forms of participation in online communities (Fan and Zhang, 2020; Zhang et al., 2020; Jhaver et al., 2021) and how the status quo designs are embedded with anti-democratic ideologies that are revealed in things like permission structures, making it especially challenging for change (Massanari, 2017; Schneider, 2022). In order to understand how social arrangements emerge and become reproduced through technology design, I consider the call by Leonardi and Barley (2008) “to capture variation in socio-material configurations and opportunities for change” (p. 166). In other words, we have to compare radically different technologies in the same or similar contexts to understand how and why governance outcomes vary across different sociotechnical systems.

Considering the materiality of software that creates the spaces for online communities to exist and thrive, we can identify affordances and design choices that facilitate the evolution of online community governance. Affordances, again, are the possibilities for action embedded in a technology that “enables or constrains potential behavioral outcomes in a particular context” (Evans et al., 2017, p. 36). From this limited selection of experiences from leaders on Reddit, it seems as if subreddit leaders experiment more with change in governance and opening up democratic forms of participation to their community members. One possible reason for this is because community member voice has relatively high visibility compared to member voice on Fandom and Fediverse communities. Reddit communities are centered on posts and comments within posts and community members can collectively push these posts to the top of a subreddit’s page. For these reasons, members can voice concerns about governance in their community and potentially have their message seen by a much larger audience than they would see on Fandom or Fediverse communities. While members of Fandom communities can voice concerns about deterioration, they’re usually sequestered on specific article’s Talk Pages or in less frequent cases, show up on the wiki’s Discussions forum or alternatively in a Discord server where members chat. Fediverse instances have no central forum for discussing problems in their communities but they can show up on someone’s feed. However, the severity of the problem being called out might require a massive number of reposts for the message to carry any weight that would move community administrators. In both Fandom and Fediverse communities, these discussions are less

centralized and less visible than they are on Reddit. Meta discussions about the state of a community's governance being insufficient or inadequate in handling problems, when highly visible to other community members, may reinforce the problem in leaders and require some form of action and potentially change.

Chapter 8

SYNTHESIS AND CONCLUSION

8.1 Introduction

This chapter brings together the findings from Chapters 5, 6, and 7 to synthesize understandings about online community rules, their enforcement, and their evolution. I first summarize those findings and synthesize from our analysis the theoretical and empirical contributions that help understand the questions that motivated this study: *why and how do leaders of online communities adopt, enforce, and change their rules?* I then describe a set of design recommendations and implications for volunteer leaders of online communities. This chapter ends on a brief discussion of the limitations of the study, future research, and concluding thoughts.

8.2 Synthesis of Findings and Contributions

Chapter 5 investigated why and how online leaders of online communities in our study adopted their rules. We found that rules were initially adopted with the intention to regulate the behavior and moderate the content of member contributions to their communities. The rules of the communities in our study (as reflected in Chapter 3) did, after all, focus mostly on aspects of member behavior, like being civil in interactions, and on sharing relevant information. Incivility between members and low quality information sharing are the primary problems solved through content moderation in online communities (Grimmelmann, 2017).

That said, our analysis showed how rules are critical in establishing and reinforcing community boundaries, helping to preserve their identity in the face of pressures and shifts in membership. This was a problem for communities that experienced shifts in their membership due to their community being discoverable on a large platform and where participation was open to all. Influxes of newcomers can shift the demographic makeup of a community,

which may threaten its goal of specializing in a topic where not many people have expertise. Rules can help rein in on maintaining the boundaries of the community's identity by prohibiting content that doesn't align with the community's specific niche interest. While Butler et al. (2008) suggested that rules shape community identity, our findings reveal a more complex motivation: communities may be driven to implement formal rules not just to create identity, but as a defensive mechanism to preserve it, particularly through strict participation and contribution guidelines. In line with Hwang and Foote's (2021) observation that communities thrive by maintaining an optimal size (including small), our research shows how overlapping communities in topics and membership contribute to a broader ecosystem of niche groups, with some specializing in member expertise (TeBlunthuis et al., 2022). Rules can then become essential for maintaining community boundaries, helping communities reinforce their identity in dense community ecosystems.

Our findings also demonstrated that horizontal, community-driven forces of conformity may emerge in decentralized networks, even without a central, hierarchical authority. In these environments, new community leaders often adopt rule sets to signal legitimacy and attract newcomers, drawing inspiration from or directly copying prominent, high status communities. This was common among leaders on the Fediverse, where rules were seen as a way to gain a form of external legitimacy, particularly in a network of interconnected communities with the power to block one another. Inadequate governance in one community can spill problems into the broader Fediverse network. Adopting rules from high-status, legitimate communities like **Mastodon.social** served as a signal of responsible governance to a perceived audience of other Fediverse leaders. Our research here uncovered strong evidence of institutional isomorphic pressures influencing the adoption of similar rules. Contrary to the assumption that decentralized systems lack accountability (Nicholson et al., 2023), our findings highlight how coercive, mimetic, and normative isomorphic pressures, as described by DiMaggio and Powell (1983), drive these communities to adopt high standards for content moderation and civility. In the absence of a central, hierarchical authority, our findings show that horizontal, community-driven forces of conformity may emerge to maintain order within decentralized community networks.

In Chapter 6, I examined how and when rules are enforced, revealing several challenges

and strategies in these processes. Leaders noted that newcomers often ignore the rules, despite proactive efforts like welcome messages, registration requirements, and visibly displayed rules. Socialization aimed at reinforcing the rules were largely ineffective. As a result, compliance was usually enforced through reactive measures like removing violations and issuing warnings. Some leaders used graduated sanctions to educate members without alienating them, while others abandoned this approach, citing time constraints or a zero-tolerance stance on serious violations like hate speech. Rule enforcement, in practice, often involved “grey areas,” where it was unclear if a rule had been broken, requiring leaders to gather evidence and deliberate with others to interpret the situation accurately. Some leaders engaged in additional practices to build legitimate authority by building trust through open communication and soliciting feedback from members.

I highlight two contributions from this chapter. First, our findings build on prior research on the labor of volunteer content moderation (Dosono and Semaan, 2019; Matias, 2019; Seering, 2019; Wohn, 2019), stressing how the additional practices of rule enforcement and building legitimacy can strain volunteer leaders’ work capacity, limiting their ability to govern effectively. Although graduated sanctions are widely regarded as an effective strategy for equitable community governance (Ostrom, 1990; Kraut et al., 2012), our findings reveal that leaders often abandoned these methods as enforcement became increasingly burdensome due to the scale and intensity of their total moderation tasks. Additionally, leaders may resort to using anonymous alternative accounts to cope with the emotional toll of content moderation, shielding their mental health from the abused aimed at their official moderator identities.

This chapter also sheds light on how online community governance operates in practice. Contrary to arguments that platforms facilitate “feudal” relationships between leaders and members (Schneider, 2022), I observed a strong influence of collective membership on leadership decisions *despite* the embedded hierarchical structures that lack accountability mechanisms for leadership. While platforms like Reddit may not guarantee members an “effective voice” in driving change (Frey and Schneider, 2021), I found that leaders valued member input for problem-solving and trust-building, which enhanced both retention and rule compliance. Moreover, encouraging member voice and compromise may be crit-

ical in addressing potential leadership challenges and avoiding regime changes driven by dissatisfaction among members.

Lastly, Chapter 7 explored the factors that drive rule changes in online communities. I found that rule changes are often triggered by persistent moderation issues, such as toxic behavior or off-topic content, which can be exacerbated by an influx of newcomers. Other triggers include community maturation, where outdated rules no longer fit modern needs, such as shifts in cultural norms or declining participation, and regime changes, where new leaders revise rules they believed contributed to the previous leadership's failure. Leaders framed rule changes using terms like “cleaning” or “fixing” to address unclear or insufficient rules. In some cases, rules remain unchanged due to a lack of noticeable problems, particularly in smaller, homogeneous communities. Leaders also cited time constraints and the voluntary nature of their work as barriers to rule changes. Finally, interviews helped us describe a general rule process: identifying problems, assessing their importance, deliberating potential solutions, and deciding whether to implement the changes.

Our analysis of why and how rules change in online communities revealed several insights contributing to existing theories of online communities. First, while prior studies of rules in online communities examines additions to communities' rule sets (Frey and Sumner, 2019; Reddy and Chandrasekharan, 2023), I found that communities in decline may remove rules to encourage more varieties of participation. This is in contrast to Frey and Sumner's (2019) statistical analysis of the evolution of rules in online communities which found that communities may succeed when their governance becomes more complex and Reddy and Chandrasekharan's (2023) work that found communities usually add rules over time that restrict kinds of content. In both of those studies, the assumption is on community growth as a catalyst for change in governance, but our work highlights an often overlooked aspect of online community life-cycles: decline. Leaders may have to make compromises and broaden the acceptable forms of participation to keep the community active as it faces diminishing engagement. Our findings suggest that successful governance isn't solely about increasing rule complexity or restriction but also about adapting to the realities of their environment.

I also extend work by (Frey and Schneider, 2021) by showing how the materiality and affordances of different sociotechnical designs can facilitate or constrain the presence of “ef-

fective voice” in online community governance and create opportunities for the evolution of community governance. In the absence of platform-provided mechanisms for transparent and binding decision-making processes, many online communities fail to include meaningful member participation in governance. However, I found that on Reddit, the combination of visible and centralized discussion spaces and voting mechanisms enabled opportunities for effective voice in some communities. Discussions are less visible on Fandom wikis, where one has to search out the Talk Pages on specific articles to become aware of conflicts at all. Leaders would consider member’s opinions and feedback in some instances but interviews with Fandom leaders suggested that the division of labor often kept decision-making concentrated among a small group, limiting broader community involvement. While member voice on the Fediverse is largely more visible than it is on Fandom, the structure of Fediverse instances can lead to homophilous communities. Membership requires a registration process and users appear to self-select into communities whose rules already align with their own preferences for online community experiences. In sum, our cases show how the design and affordances of different platforms can “occasion” (Barley, 1986) opportunities for community governance evolution. The lack of platform-provided mechanisms for leadership accountability does not necessarily mean that community-driven change won’t occur, but I argue that some technical designs may facilitate the opportunities for these changes better than others.

8.3 *Implications for Design and Leadership*

The lessons drawn from my analysis of the findings in this research can inform future design choices for platforms that support online communities. First, designers of community platforms may want to create rule set templates from which leaders may select and customize as their own. In interviews, leaders described spending time doing research on other communities’ rules to model as their own, but the cost of searching may lead to leaders to “satisficing”, in which they adopt rules that are adequate enough but perhaps not optimal (Cyert and March, 1963). In Chapter 5, I showed how initial rule adoption may be influenced by institutional isomorphic pressures to mimic communities that are perceived as more successful, and this influence could result in leaders choosing rules that may not

align as well with their own goals. A system for helping leaders create custom rule sets for their specific needs could help avoid this type of mimetic organizational behavior. We could, for example, imagine during initial community creation, leaders may be prompted to select a rule set based on a set of parameters such as the goals (e.g., discuss a topic, share original content, trade goods and services, etc.), and behavioral or discussion norms they'd like to have, and these parameters could be used to generate templates for further customization. Rule language text could be automatically generated with the assistance of large language models (LLMs) to help leaders with little experience in writing these kinds of formal regulatory documents.

Designers might also consider additional tools and systems for helping volunteer leaders of online communities mediate disputes between leaders and members in the rule enforcement process. Interviews with leaders about rule enforcement revealed complex practices of evidence collection and context gathering of rule violators. Some leaders found this work too exhausting to keep up with given the scale of their work in a very large and active community, and they would essentially forego these practices and enforce the highest sanctions on rule violators instead of offering opportunities to save face. While Kraut et al. (2012) suggest leaders do this work because it helps build trust in community governance, it often doesn't scale well with larger communities or communities whose leadership teams are understaffed. Work by Cai and Wohn (2021) offer one recommendation suggesting platform designers build mechanisms for collaborative profiling of rule violators on Twitch to help in these cases. This would allow sharing of information about rule violators on a platform to more quickly understand that user's history and have a clearer understanding of the situation they're in. That said, I offer another design recommendation in which context might be also gathered through an automated system that generates a story of events leading up to the rule violation. Resolving the issue of mediating disputes at a higher scale of member activity, the recommendation might be as simple as making it easier for volunteer leaders to recruit and train new content moderators to assist with these tasks in their communities. Educational materials and training programs created by the platform could help onboard new leaders quickly into their roles. We might also imagine implementing the automated suggestion to community leaders to recruit more volunteers if their community is becoming

more active or increasing in size past the rate in which everyday governance can be fairly carried out.

Another design recommendation based on the lessons of this research is to consider in community creation the intention of specific governance structures and offer alternative templates. While some leaders evolved the governance of their communities to include participatory democratic processes in the form of polling, deliberating about problems, and voting on governance decisions, this was still not a widespread occurrence among leaders interviewed. This recommendation echoes Frey and Schneider’s (2021) call for more “effective voice” mechanisms in the design of platforms that host online communities, but going further, I argue that designers should include in community creation alternative forms of governing. While it’s true that the hierarchical structure embedded into these platforms can result in seemingly oligarchical or anti-democratic control over communities (Schneider, 2022), leaders can and do experiment beyond the limitations of the platform’s initial design with democratic practices. However, these are often improvised solutions and at the discretion of leaders based on their preferences for governing. For example, platform designers could include templates for governance based on the ideas in Community Rule¹ system created by University of Colorado Boulder’s Media Economies Design Lab. Leaders may consider an alternative systems of decision-making for their communities, like a consensus based governance model or a council of representatives to act on behalf of members in decisions made by community leaders. Such alternative models open up the possibilities for solving problems that facilitate the community’s growth, like having the members decide if a contentious rule should be implemented or not through a voting process. Another example might be models that include term limits could be useful for volunteer leaders that are burned out from their work or have failed the community.

Finally, I recommend designers consider mechanisms for alerting leaders about adapting their rules to a different environment. Reddy and Chandrasekharan (2023) made this recommendation in the case of Reddit but under the assumption that it was through growth that a community might need to be alerted to alter or change its rules. I recommend that

¹<https://communityrule.info/templates/>

such systems might also alert leaders when communities are in periods of decline. Leaders who found their rules to be too restricting considered removing those restrictive rules to promote more activity and deter terminal decline. Additionally, prompting leaders to change the rules should be accompanied with the creation of a space for members to give feedback about possible rule change solutions. This information can help inform the correct decisions to make and build trust between leaders and members. While Bulat et al.'s (2024) survey study found that moderators who employ participatory approaches to community governance feel less community acceptance and belonging, I show that such practices have benefits for community stability and can in some cases reinforce leaders' positions of authority. Designing for participation of community members in governance decision making during periods of decline or crises may be beneficial in successfully navigating out of these dilemmas and stabilizing membership.

8.4 *Limitations and Future Research*

We sought out to interview a large portion of experienced leaders from different online communities and sample from distinct sociotechnical systems to achieve sufficient but varied comparative case studies in this dissertation. Interviews were open-ended and responses often involved deep reflections on leaders' experiences in governing their communities. The analysis of the interview data was structured and inspired by practices from Charmaz (2015), and analytical memos were iterated on repeatedly with peer-review from other researchers and scholars in this field. That said, the key findings of this work have limitations and could be improved upon through efforts in future research.

First, our data are limited to leaders from just Reddit, Fandom, and Fediverse instances. While these are distinct case studies in terms of how and why communities form on them, future work might consider raising similar questions about rule adoption and change in alternative settings. For example, chat-based communities have grown in popularity on platforms like Discord and Slack, and both are built around seemingly different goals (gaming and fandom in Discord and typically professional or organizational purposes in Slack). In both settings, discovering other communities' rules on those platforms may not be as easy as browsing subreddits, wikis, or Fediverse instances as they often require private invi-

tation links. We were also limited to the 40 interviews with leaders from the three settings we did select to include in this study. With only nine from Reddit, 14 from Fandom, and 17 from the Fediverse, some may argue that we could have interviewed a larger number of leaders from each setting, especially on Reddit. Coincidentally, Reddit has been virtually at the forefront of empirical case studies on volunteer community governance research in some of the most visible publication venues for this field since its creation in 2005, and our section in Chapter 3 on the empirical background of Reddit references a tremendous amount of research on the leaders and rules of their communities. While research specific to the Fandom platform is limited in scope, the body of research on communities built around wiki infrastructure, created in 1995, surpass even that of research on Reddit. Research on the Fediverse is comparatively much less common, perhaps because it's still relatively fresh off its nascent stage with the introduction of Mastodon in 2016. That said, even if we had a potentially larger sample, we reached saturation in interview responses before ending participant recruitment with the current sample size.

Some may also argue that I can't generalize our findings from interview data alone. However, it is not the goal of interpretive research methods such as the ones used in this study to achieve external validity. In fact, it is in the unique experiences from our cases that I may draw conclusions about the *possibilities* of social realities and organizational practices. That said, future research could be informed by our findings and statistically test some of our propositions about how similar rules get adopted and when they're likely to change. Researchers like Frey and Sumner (2019) and Reddy and Chandrasekharan (2023) have statistically analyzed the relationship between factors like community maturity and community membership size to predict occurrences for changes in the rules. I found that change could be limited by the capacity of leaders to find time to respond to problems with a change in the rules. I also found that a decrease in community member activity might also occasion the opportunity for rule change. Statistical tests could incorporate the potential correlate of leadership team size as well as membership decline in predicting a change in the rules.

8.5 *Conclusion*

In a time where online communities have become salient across modern social life, understanding how and why the volunteers who govern them create, enforce, and change policies and rules is crucial for sustaining their longevity. They are not just spaces for online entertainment; online communities have become vital to the peer production of knowledge and the facilitation of cultural and political discourse. At the same time, they exist within fragile ecosystems, often under threat from waves of new members, bad actors seeking to disrupt, and the burnout or mistakes of the very leaders who sustain them. As my research shows, the governance of online communities is not a purely mechanical process—it is deeply human, shaped by the choices, constraints, and evolving strategies of volunteer leaders who work to balance the competing demands of enforcing rules on and upholding the social contract with their members. Without strategic governance of their spaces, online communities risk deteriorating under external pressures and internal strain. It is imperative that designers of community platforms and social computing scholars recognize the critical role that governance plays in shaping the success or failure of online communities and ensure that the systems we build empower their leaders to adapt and thrive in an ever-changing sociotechnical landscape.

BIBLIOGRAPHY

- Ammari, T., Nofal, M., Naseem, M., and Mustafa, M. (2022). Moderation as Empowerment: Creating and Managing Women-Only Digital Safe Spaces. *Proc. ACM Hum.-Comput. Interact.*, 6(CSCW2):313:1–313:36.
- Anaobi, I. H., Raman, A., Castro, I., Zia, H. B., Ibosiola, D., and Tyson, G. (2023). Will Admins Cope? Decentralized Moderation in the Fediverse. In *Proceedings of the ACM Web Conference 2023*, WWW '23, pages 3109–3120, New York, NY, USA. Association for Computing Machinery.
- Arazy, O., Lindberg, A., Rezaei, M., and Samorani, M. (2019). The evolutionary trajectories of peer-produced artifacts: Group composition, the trajectories' exploration, and the quality of artifacts. *MIS Quarterly*.
- Arnold, T., Daxenberger, J., Gurevych, I., and Weihe, K. (2017). Is Interaction More Important Than Individual Performance?: A Study of Motifs in Wikia. In *Proceedings of the 26th International Conference on World Wide Web Companion*, WWW '17 Companion, pages 1609–1617, Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee.
- Atreja, S., Im, J., Resnick, P., and Hemphill, L. (2024). AppealMod: Inducing Friction to Reduce Moderator Workload of Handling User Appeals. *Proc. ACM Hum.-Comput. Interact.*, 8(CSCW1):19:1–19:35.
- Barbosa, S., Cosley, D., Sharma, A., and Cesar, Jr., R. M. (2016). Averaging gone wrong: Using time-aware analyses to better understand behavior. In *Proceedings of the 25th International Conference on World Wide Web (WWW '16)*, pages 829–841, Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee.

- Barley, S. R. (1986). Technology as an occasion for structuring: evidence from observations of ct scanners and the social order of radiology departments. *Administrative Science Quarterly*, 31(1):78–108.
- Baum, J. A. C. and Shipilov, A. V. (2006). Ecological approaches to organizations. In *Sage Handbook for Organization Studies*, pages 55–110. Sage, Rochester, NY.
- Bellah, R. N., Madsen, R., Sullivan, W. M., Swidler, A., and Tipton, S. M. (1985). *Habits of the heart: individualism and commitment in American life*. University of California Press, Berkeley, Calif. Los Angeles, Calif. London. OCLC: 845177699.
- Berger, P. L. and Luckmann, T. (1966). *The social construction of reality: a treatise in the sociology of knowledge*. Anchor Books, New York, repr edition.
- Bernstein, M. S., Monroy-Hernández, A., Harry, D., André, P., Panovich, K., and Vargas, G. (2011). 4chan and /b/: An analysis of anonymity and ephemerality in a large online community. In *Fifth International AAAI Conference on Weblogs and Social Media*, Palo Alto, CA. AAAI Publications. tex.ids= bernstein_4chan_2011-1 number: 1.
- Bevir, M. (2012). *Governance: A very short introduction*. Number 333 in Very short introductions. Oxford University Press, Oxford, United Kingdom, first edition edition. OCLC: ocn797975142.
- Bijker, W. E. (2010). How is technology made?—That is the question! *Cambridge Journal of Economics*, 34(1):63–76. Publisher: Oxford Academic.
- Billings, M. and Watts, L. A. (2010). Understanding Dispute Resolution Online: Using Text to Reflect Personal and Substantive Issues in Conflict. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 1447–1456, New York, NY, USA. ACM.
- Bipat, T., Fussell, S. R., Hecht, B., Kiene, C., McDonald, D. W., and Zachry, M. (2018). Navigating the challenges of multi-site research. In *Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '18 Companion, pages 409–415, New York, NY, USA. Association for Computing Machinery.

- Black, L. W., Welser, T., DeGroot, J. M., and Cosley, D. (2008). "Wikipedia is not a democracy": deliberation and policy-making in an online community.
- Boczkowski, P. J. (1999). Mutual shaping of users and technologies in a national virtual community. *Journal of Communication*, 49(2):86–108. Publisher: Wiley-Blackwell.
- Bono, C. A., La Cava, L., Luceri, L., and Pierri, F. (2024). An Exploration of Decentralized Moderation on Mastodon. In *Proceedings of the 16th ACM Web Science Conference, WEBSCI '24*, pages 53–58, New York, NY, USA. Association for Computing Machinery.
- Bourne, C. D. (2017). Sensemaking in an online community after financial loss: Enterprising Jamaican investors and the fall of a financial messiah. *New Media & Society*, 19(6):843–860. Publisher: SAGE Publications.
- Boyd, D. M. and Ellison, N. B. (2007a). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, 13(1):210–230. Publisher: Oxford Academic.
- Boyd, D. M. and Ellison, N. B. (2007b). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, 13(1):210–230. Publisher: Oxford Academic.
- Bruckman, A. (2006). A new perspective on "community" and its implications for computer-mediated communication systems. In *CHI '06 extended abstracts on Human factors in computing systems*, pages 616–621, Montréal, Québec, Canada. ACM.
- Bulat, B., Wang, H., Fujimoto, S., and Frey, S. (2024). The psychology of volunteer moderators: Tradeoffs between participation, belonging, and norms in online community governance. *New Media & Society*, page 14614448241259028. Publisher: SAGE Publications.
- Butler, B. S. (2001). Membership size, communication activity, and sustainability: A resource-based model of online social structures. *Information Systems Research*, 12(4):346–362.

- Butler, B. S., Joyce, E., and Pike, J. (2008). Don't look now, but we've created a bureaucracy: the nature and roles of policies and rules in Wikipedia. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 1101–1110, New York, NY, USA. ACM.
- Caelin, D. (2022). Decentralized Networks vs The Trolls. In Mahmoudi, H., Allen, M. H., and Seaman, K., editors, *Fundamental Challenges to Global Peace and Security : The Future of Humanity*, pages 143–168. Springer International Publishing, Cham.
- Cai, J. and Wohn, D. Y. (2019). What are Effective Strategies of Handling Harassment on Twitch? Users' Perspectives. In *Companion Publication of the 2019 Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '19 Companion, pages 166–170, New York, NY, USA. Association for Computing Machinery.
- Cai, J. and Wohn, D. Y. (2021). After Violation But Before Sanction: Understanding Volunteer Moderators' Profiling Processes Toward Violators in Live Streaming Communities. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW2):410:1–410:25.
- Campbell, J., Fletcher, G., and Greenhill, A. (2009). Conflict and identity shape shifting in an online financial community. *Information Systems Journal*, 19(5):461–478.
- Carr, C. T. (2020). CMC Is Dead, Long Live CMC!: Situating Computer-Mediated Communication Scholarship Beyond the Digital Age. *Journal of Computer-Mediated Communication*, 25(1):9–22. Publisher: Oxford Academic.
- Carr, C. T. and Hayes, R. A. (2015). Social Media: Defining, Developing, and Divining. *Atlantic Journal of Communication*, 23(1):46–65.
- Carroll, G. R. and Hannan, M. T. (1989). Density dependence in the evolution of populations of newspaper organizations. *American Sociological Review*, 54(4):524.
- Chakraborti, M., Bulat, B., Zhong, Q., Sen, A., and Frey, S. (2022). Deconstructing written rules and hierarchy in peer produced software communities. arXiv:2206.07992 [cs].

- Chancellor, S., Hu, A., and De Choudhury, M. (2018). Norms Matter: Contrasting Social Support Around Behavior Change in Online Weight Loss Communities. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 1–14, Montreal QC, Canada. Association for Computing Machinery.
- Chancellor, S., Pater, J. A., Clear, T., Gilbert, E., and De Choudhury, M. (2016). #thygh-gapp: Instagram Content Moderation and Lexical Variation in Pro-Eating Disorder Communities. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, CSCW '16, pages 1201–1213, New York, NY, USA. Association for Computing Machinery.
- Chandrasekharan, E., Gandhi, C., Mustelier, M. W., and Gilbert, E. (2019). Crossmod: a cross-community learning-based system to assist reddit moderators. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–30.
- Chandrasekharan, E., Pavalanathan, U., Srinivasan, A., Glynn, A., Eisenstein, J., and Gilbert, E. (2017). You can't stay here: The efficacy of reddit's 2015 ban examined through hate speech. *Proc. ACM Hum.-Comput. Interact.*, 1(CSCW):31:1–31:22. Citation Key Alias: chandrasekharan_you.2017-1.
- Chandrasekharan, E., Samory, M., Jhaver, S., Charvat, H., Bruckman, A., Lampe, C., Eisenstein, J., and Gilbert, E. (2018). The Internet's hidden rules: An empirical study of Reddit norm violations at micro, meso, and macro scales. *Proc. ACM Hum.-Comput. Interact.*, 2(CSCW):32:1–32:25.
- Charmaz, K. (2015). *Constructing grounded theory: A practical guide through qualitative analysis*. SAGE, Thousand Oaks, California, 2nd edition. tex.ids=charmaz_constructing_2006, charmaz_constructing.2006-1, charmaz_constructing.2006-2, charmaz_constructing.2006-3, charmaz_constructing.2014 lccn: H61.24.C45 2006.
- Chatman, J. A. (1991). Matching People and Organizations: Selection and Socialization in Public Accounting Firms. *Administrative Science Quarterly*, 36(3):459–484. Publisher: [Sage Publications, Inc., Johnson Graduate School of Management, Cornell University].

- Coleman, G. and Hill, B. M. (2004). The social production of ethics in Debian and free software communities: anthropological lessons for vocational ethics. In Koch, S., editor, *Free/Open Source Software Development*. Idea Group Inc. (IGI), Hershey, Pennsylvania.
- Coleman, J. S. (1988). Social Capital in the Creation of Human Capital. *American Journal of Sociology*, 94:S95–S120.
- Colglazier, C., TeBlunthuis, N., and Shaw, A. (2024). The Effects of Group Sanctions on Participation and Toxicity: Quasi-experimental Evidence from the Fediverse. *Proceedings of the International AAAI Conference on Web and Social Media*, 18:315–328.
- Cullen, A. L. L. and Kairam, S. R. (2022). Practicing Moderation: Community Moderation as Reflective Practice. *Proc. ACM Hum.-Comput. Interact.*, 6(CSCW1):111:1–111:32.
- Cyert, R. M. and March, J. G. (1963). *A behavioral theory of the firm*. Prentice-Hall international series in management. Prentice-Hall, Englewood Cliffs, N.J.
- Dabbish, L., Farzan, R., Kraut, R., and Postmes, T. (2012). Fresh faces in the crowd: turnover, identity, and commitment in online groups. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work, CSCW '12*, pages 245–248, New York, NY, USA. Association for Computing Machinery.
- De Filippi, P. and Loveluck, B. (2016). The Invisible Politics of Bitcoin: Governance Crisis of a Decentralized Infrastructure.
- DeVito, M. A., Walker, A. M., and Fernandez, J. R. (2021). Values (Mis)alignment: Exploring Tensions Between Platform and LGBTQ+ Community Design Values. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW1):88:1–88:27.
- DiMaggio, P. J. and Powell, W. W. (1983). The iron cage revisited: institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, 48(2):147–160.
- Doehne, M., Geweke, J., and Rost, K. (2023). Aleatoric governance: using lotteries to break the iron law of oligarchy. *European Sociological Review*, 39(4):646–662.

- Dosono, B. and Semaan, B. (2019). Moderation practices as emotional labor in sustaining online communities: the case of aapi identity work on reddit. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 1–13, Glasgow, Scotland Uk. Association for Computing Machinery. `tex.ids= dosono_moderation.2019`.
- Ducheneaut, N., Yee, N., Nickell, E., and Moore, R. J. (2007). The life and death of online gaming communities: a look at guilds in world of warcraft. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, pages 839–848, New York, NY, USA. ACM.
- Dunbar-Hester, C. (2024). Showing your ass on Mastodon: Lossy distribution, hashtag activism, and public scrutiny on federated, feral social media. *First Monday*.
- Evans, S. K., Pearce, K. E., Vitak, J., and Treem, J. W. (2017). Explicating affordances: a conceptual framework for understanding affordances. *Journal of Computer-Mediated Communication*, 22(1):35–52.
- Fan, J. and Zhang, A. X. (2020). Digital juries: a civics-oriented approach to platform governance. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, pages 1–14, New York, NY, USA. Association for Computing Machinery.
- Farzan, R., Dabbish, L. A., Kraut, R. E., and Postmes, T. (2011). Increasing commitment to online communities by designing for social presence. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, CSCW '11, pages 321–330, New York, NY, USA. Association for Computing Machinery.
- Fernback, J. (2007). Beyond the diluted community concept: a symbolic interactionist perspective on online social relations. *New Media & Society*, 9(1):49–69. Publisher: SAGE Publications.
- Fiesler, C., Jiang, J. A., McCann, J., Frye, K., and Brubaker, J. R. (2018). Reddit rules! Characterizing an ecosystem of governance. In *Proceedings of the International AAAI*

- Conference on Web and Social Media*, volume 12, pages 72–81, Stanford, CA. AAAI. tex.ids= flesler_reddit_2018-1 number: 1.
- Fiesler, C., Lampe, C., and Bruckman, A. S. (2016). Reality and Perception of Copyright Terms of Service for Online Content Creation. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, CSCW '16, pages 1450–1461, New York, NY, USA. Association for Computing Machinery.
- Foot, K. A. (2014). Cultural-historical activity theory: exploring a theory to inform practice and research. *Journal of Human Behavior in the Social Environment*, 24(3):329–347.
- Foote, J. and Contractor, N. (2018). The behavior and network position of peer production founders. In Chowdhury, G., McLeod, J., Gillet, V., and Willett, P., editors, *iConference 2018: Transforming Digital Worlds*, Lecture Notes in Computer Science, pages 99–106. Springer.
- Foote, J., Gergle, D., and Shaw, A. (2017). Starting online communities: motivations and goals of wiki founders. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*, pages 6376–6380, New York, NY. ACM.
- Freeman, J. and Hannan, M. T. (1983). Niche width and the dynamics of organizational populations. *American Journal of Sociology*, 88(6):1116–1145.
- Frey, S., Krafft, P. M., and Keegan, B. C. (2019). "This Place Does What It Was Built For": Designing Digital Institutions for Participatory Change. *Proc. ACM Hum.-Comput. Interact.*, 3(CSCW):32:1–32:31.
- Frey, S. and Schneider, N. (2021). Effective voice: Beyond exit and affect in online communities. *New Media & Society*, page 14614448211044025. Publisher: SAGE Publications.
- Frey, S. and Sumner, R. W. (2018). Emergence of complex institutions in a large population of self-governing communities. *arXiv:1804.10312 [cs]*. arXiv: 1804.10312.
- Frey, S. and Sumner, R. W. (2019). Emergence of integrated institutions in a large pop-

- ulation of self-governing communities. *PLOS ONE*, 14(7):e0216335. Publisher: Public Library of Science.
- Frey, S., Zhong, Q., Bulat, B., Weisman, W. D., Liu, C., Fujimoto, S., Wang, H., and Schweik, C. M. (2022). Governing Online Goods: Maturity and Formalization in Minecraft, Reddit, and World of Warcraft Communities. *Proc. ACM Hum.-Comput. Interact.*, 6(CSCW2):300:1–300:23.
- Gaver, W. W. (1991). Technology Affordances. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '91, pages 79–84, New York, NY, USA. ACM. event-place: New Orleans, Louisiana, USA.
- Gaver, W. W. (1992). The Affordances of Media Spaces for Collaboration. In *Proceedings of the 1992 ACM Conference on Computer-supported Cooperative Work*, CSCW '92, pages 17–24, New York, NY, USA. ACM. event-place: Toronto, Ontario, Canada.
- Gehl, R. W. and Zulli, D. (2023). The digital covenant: non-centralized platform governance on the mastodon social network. *Information, Communication & Society*, 26(16):3275–3291. Publisher: Routledge _eprint: <https://doi.org/10.1080/1369118X.2022.2147400>.
- Geiger, R. S. (2014). Bots, bespoke, code and the materiality of software platforms. *Information, Communication & Society*, 17(3):342–356.
- Geiger, R. S. and Ford, H. (2011). Participation in Wikipedia’s Article Deletion Processes. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration*, WikiSym '11, pages 201–202, New York, NY, USA. ACM.
- Geiger, R. S. and Ribes, D. (2010). The work of sustaining order in wikipedia: the banning of a vandal. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*, pages 117–126, New York, NY, USA. Association for Computing Machinery.
- Giddens, A. (1984). *The constitution of society: outline of the theory of structuration*. University of California Press. Google-Books-ID: x2bf4g9Z6ZwC.

- Gilbert, E. (2013). Widespread underprovision on Reddit. In *Proceedings of the 2013 conference on Computer supported cooperative work, CSCW '13*, pages 803–808, San Antonio, Texas, USA. Association for Computing Machinery.
- Gilbert, S. (2023). Towards Intersectional Moderation: An Alternative Model of Moderation Built on Care and Power. *Proc. ACM Hum.-Comput. Interact.*, 7(CSCW2):256:1–256:32.
- Gilbert, S. A. (2020). "I run the world's largest historical outreach project and it's on a cesspool of a website." Moderating a Public Scholarship Site on Reddit: A Case Study of r/AskHistorians. *Proc. ACM Hum.-Comput. Interact.*, 4(CSCW1):19:1–19:27.
- Gillespie, T. (2018). *Custodians of the internet: platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press, New Haven. OCLC: on1005113962.
- Grimmelmann, J. (2017). The Virtues of Moderation. preprint, LawArXiv.
- Halfaker, A., Geiger, R. S., Morgan, J. T., and Riedl, J. (2013). The rise and decline of an open collaboration system: how Wikipedia's reaction to popularity is causing its decline. *American Behavioral Scientist*, 57(5):664–688.
- Halfaker, A., Kittur, A., and Riedl, J. (2011). Don't bite the newbies: How reverts affect the quantity and quality of Wikipedia work. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration (WikiSym '11)*, pages 163–172, New York, NY. ACM.
- Hannan, M. T. and Freeman, J. (1977). The population ecology of organizations. *American Journal of Sociology*, 82(5):929–964.
- Hill, B. M. and Shaw, A. (2021). The hidden costs of requiring accounts: Quasi-experimental evidence from peer production. *Communication Research*, 48(6):771–795.
- Hillman, A. J., Withers, M. C., and Collins, B. J. (2009). Resource Dependence Theory: A Review. *Journal of Management*, 35(6):1404–1427. Publisher: SAGE Publications Inc.

- Hirschman, A. O. (1970). *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States*. Harvard University Press.
- Horta Ribeiro, M., Jhaver, S., Zannettou, S., Blackburn, J., Stringhini, G., De Cristofaro, E., and West, R. (2021). Do Platform Migrations Compromise Content Moderation? Evidence from r/The_donald and r/Incels. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW2):316:1–316:24.
- Hwang, S. and Foote, J. D. (2021). Why do people participate in small online communities? *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2):1–25.
- Hwang, S. and Shaw, A. (2022). Rules and Rule-Making in the Five Largest Wikipedias. *Proceedings of the International AAAI Conference on Web and Social Media*, 16:347–357.
- Iriberri, A. and Leroy, G. (2009). A life-cycle perspective on online community success. *ACM Computing Surveys*, 41(2):1–29.
- Jhaver, S., Appling, D. S., Gilbert, E., and Bruckman, A. (2019a). "Did You Suspect the Post Would be Removed?": Understanding User Reactions to Content Removals on Reddit. *Proc. ACM Hum.-Comput. Interact.*, 3(CSCW):192:1–192:33.
- Jhaver, S., Birman, I., Gilbert, E., and Bruckman, A. (2019b). Human-machine collaboration for content regulation: the case of reddit automoderator. *ACM Trans. Comput.-Hum. Interact.*, 26(5):31:1–31:35.
- Jhaver, S., Bruckman, A., and Gilbert, E. (2019c). Does transparency in moderation really matter? User behavior after content removal explanations on reddit. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):150:1–150:27.
- Jhaver, S., Frey, S., and Zhang, A. (2021). Designing for multiple centers of power: a taxonomy of multi-level governance in online social platforms. arXiv: 2108.12529.
- Jhaver, S., Frey, S., and Zhang, A. X. (2023a). Decentralizing Platform Power: A Design Space of Multi-Level Governance in Online Social Platforms. *Social Media + Society*, 9(4):20563051231207857. Publisher: SAGE Publications Ltd.

- Jhaver, S., Zhang, A. Q., Chen, Q. Z., Natarajan, N., Wang, R., and Zhang, A. X. (2023b). Personalizing Content Moderation on Social Media: User Perspectives on Moderation Choices, Interface Design, and Labor. *Proc. ACM Hum.-Comput. Interact.*, 7(CSCW2):289:1–289:33.
- Jhaver, S. and Zhang, A. X. (2023). Do users want platform moderation or individual control? Examining the role of third-person effects and free speech support in shaping moderation preferences. *New Media & Society*, page 14614448231217993. Publisher: SAGE Publications.
- Jiang, J. A., Kiene, C., Middler, S., Brubaker, J. R., and Fiesler, C. (2019). Moderation challenges in voice-based online communities on discord. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):23.
- Jiang, J. A., Middler, S., Brubaker, J. R., and Fiesler, C. (2020). Characterizing Community Guidelines on Social Media Platforms. In *Companion Publication of the 2020 Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '20 Companion, pages 287–291, New York, NY, USA. Association for Computing Machinery.
- Jimenez-Diaz, G., Serrano, A., and Arroyo, J. (2018). A Wikia census: motives, tools and insights. In *Proceedings of the 14th International Symposium on Open Collaboration*, OpenSym '18, pages 1–6, New York, NY, USA. Association for Computing Machinery.
- Jones, Q. (1997). Virtual-Communities, Virtual Settlements & Cyber-Archaeology: a Theoretical Outline. *Journal of Computer-Mediated Communication*, 3(3). Publisher: Oxford Academic.
- Jones, R., Colusso, L., Reinecke, K., and Hsieh, G. (2019). r/science: Challenges and Opportunities in Online Science Communication. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 1–14, Glasgow, Scotland Uk. Association for Computing Machinery.
- Juneja, P., Rama Subramanian, D., and Mitra, T. (2020). Through the Looking Glass:

- Study of Transparency in Reddit’s Moderation Practices. *Proc. ACM Hum.-Comput. Interact.*, 4(GROUP):17:1–17:35.
- Kairam, S. R., Wang, D. J., and Leskovec, J. (2012). The life and death of online groups: predicting group growth and longevity. In *Proceedings of the fifth ACM international conference on Web search and data mining*, WSDM ’12, pages 673–682, New York, NY, USA. Association for Computing Machinery. tex.ids= kairam_life.2012.
- Kappeler, V. E. and Schaefer, B. P. (2018). *The Police and Society: Touchstone Readings, Fourth Edition*. Waveland Press. Google-Books-ID: NB1tDwAAQBAJ.
- Kaptelinin, V. and Nardi, B. (2007). Acting with technology: activity theory and interaction design. *First Monday*, 12(4).
- Kayhan, V. O. (2015). The nature, dimensionality, and effects of perceptions of community governance. *Information & Management*, 52(1):18–29.
- Keegan, B. and Fiesler, C. (2017). The evolution and consequences of peer producing wikipedia’s rules. In *Proceedings of the Eleventh International AAAI Conference on Web and Social Media*, pages 112–121. AAAI. tex.ids= keegan_evolution.2017.
- Kharazian, Z., Starbird, K., and Hill, B. M. (2024). Governance Capture in a Self-Governing Community: A Qualitative Comparison of the Croatian, Serbian, Bosnian, and Serbo-Croatian Wikipedias. *Proc. ACM Hum.-Comput. Interact.*, 8(CSCW1):61:1–61:26.
- Kiene, C., Grandprey-Shores, K., Chandrasekharan, E., Jhaver, S., Jiang, J. A., Dym, B., Seering, J., Gilbert, S., Lo, K., Wohn, D. Y., and Dosono, B. (2019a). Volunteer work: mapping the future of moderation research. In *Companion Publication of the 2019 Conference on Computer Supported Cooperative Work and Social Computing*, CSCW ’19 Companion, pages 492–497, New York, NY, USA. Association for Computing Machinery.
- Kiene, C. and Hill, B. M. (2020). Who uses bots? A statistical analysis of bot usage in moderation teams. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI EA ’20, pages 1–8, Honolulu, HI, USA. Association for Computing Machinery. tex.ids: kiene_who.2020-1.

- Kiene, C., Jiang, J. A., and Hill, B. M. (2019b). Technological frames and user innovation: exploring technological change in community moderation teams. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):44:1–44:23.
- Kiene, C., Monroy-Hernández, A., and Hill, B. M. (2016). Surviving an “eternal september”: how an online community managed a surge of newcomers. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 1152–1156, New York, NY. ACM.
- Kiene, C., Shaw, A., and Hill, B. M. (2018). Managing organizational culture in online group mergers. *Proc. ACM Hum.-Comput. Interact.*, 2(CSCW):89:1–89–21.
- Kim, S., Eun, J., Seering, J., and Lee, J. (2021). Moderator Chatbot for Deliberative Discussion: Effects of Discussion Structure and Discussant Facilitation. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW1):87:1–87:26.
- Kline, R. and Pinch, T. (1996). Users as Agents of Technological Change: The Social Construction of the Automobile in the Rural United States. *Technology and Culture*, 37(4):763–795. Publisher: [The Johns Hopkins University Press, Society for the History of Technology].
- Konieczny, P. (2017). Decision making in the self-evolved collegiate court: Wikipedia’s Arbitration Committee and its implications for self-governance and judiciary in cyberspace. *International Sociology*, 32(6):755–774. Publisher: SAGE Publications Ltd.
- Kou, Y., Gui, X., Chen, Y., and Pine, K. (2017). Conspiracy Talk on Social Media: Collective Sensemaking during a Public Health Crisis. *Proceedings of the ACM on Human-Computer Interaction*, 1(CSCW):61:1–61:21.
- Kraut, R. E. and Fiore, A. T. (2014). The Role of Founders in Building Online Groups. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing, CSCW ’14*, pages 722–732, Baltimore, Maryland, USA. ACM. tex.ids=kraut_role.2014-1.

- Kraut, R. E., Resnick, P., and Kiesler, S. (2012). *Building Successful Online Communities: Evidence-based Social Design*. MIT Press, Cambridge, MA.
- Kumar, S., Hamilton, W. L., Leskovec, J., and Jurafsky, D. (2018). Community Interaction and Conflict on the Web. In *Proceedings of the 2018 World Wide Web Conference, WWW '18*, pages 933–943, Lyon, France. International World Wide Web Conferences Steering Committee. tex.ids= kumar_community_2018-1.
- Kuo, T., Hernani, A., and Grossklags, J. (2023). The Unsung Heroes of Facebook Groups Moderation: A Case Study of Moderation Practices and Tools. *Proc. ACM Hum.-Comput. Interact.*, 7(CSCW1):97:1–97:38.
- La Cava, L., Mandaglio, D., and Tagarelli, A. (2024). Polarization in Decentralized Online Social Networks. In *Proceedings of the 16th ACM Web Science Conference, WEBSCI '24*, pages 48–52, New York, NY, USA. Association for Computing Machinery.
- Lampe, C. and Johnston, E. (2005). Follow the (slash) dot: effects of feedback on new members in an online community. In *Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work*, pages 11–20, Sanibel Island, Florida, USA. ACM.
- Lampe, C. and Resnick, P. (2004). Slash(dot) and burn: distributed moderation in a large online conversation space. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '04*, pages 543–550, New York, NY, USA. Association for Computing Machinery. tex.ids: lampe_slashdot_2004.
- Latour, B. (1990). Technology is Society Made Durable. *The Sociological Review*, 38(1_suppl):103–131. Publisher: SAGE Publications Ltd.
- Latour, B. (2007). *Reassembling the social: an introduction to Actor-Network-Theory*. Clarendon lectures in management studies. Oxford Univ. Press, Oxford. OCLC: 254570772.
- Leavitt, A. and Clark, J. A. (2014). Upvoting hurricane Sandy: event-based news production processes on a social news site. In *Proceedings of the SIGCHI Conference on Human*

- Factors in Computing Systems*, CHI '14, pages 1495–1504, Toronto, Ontario, Canada. Association for Computing Machinery.
- Leavitt, A. and Robinson, J. J. (2017). The role of information visibility in network gate-keeping: Information aggregation on reddit during crisis events. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '17, pages 1246–1261, Portland, Oregon, USA. Association for Computing Machinery.
- Leimeister, J., Sidiras, P., and Krcmar, H. (2004). Success factors of virtual communities from the perspective of members and operators: an empirical study. *37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the*.
- Leonardi, P. M. (2012). Materiality, Sociomateriality, and Socio-Technical Systems: What Do These Terms Mean? How Are They Different? Do We Need Them? In Leonardi, P. M., Nardi, B. A., and Kallinikos, J., editors, *Materiality and Organizing: Social Interaction in a Technological World*. Oxford University Press.
- Leonardi, P. M. (2015). Studying Work Practices in Organizations: Theoretical Considerations and Empirical Guidelines. *Annals of the International Communication Association*, 39(1):235–273. Publisher: Routledge eprint: <https://doi.org/10.1080/23808985.2015.11679177>.
- Leonardi, P. M. and Barley, S. R. (2008). Materiality and change: Challenges to building better theory about technology and organizing. *Information and Organization*, 18(3):159–176. 00603.
- Lessig, L. (2006). *Code: And other laws of cyberspace, version 2.0*. Basic Books, New York, NY.
- Levi, M. (1988). *Of Rule and Revenue*. University of California Press.
- Levi, M. (2022). Trustworthy Government: The Obligations of Government & the Responsibilities of the Governed. *Daedalus*, 151(4):215–233.

- Levi, M. and Sacks, A. (2009). Legitimizing beliefs: Sources and indicators. *Regulation & Governance*, 3(4):311–333. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1748-5991.2009.01066.x>.
- Levi, M., Sacks, A., and Tyler, T. (2009). Conceptualizing Legitimacy, Measuring Legitimizing Beliefs. *American Behavioral Scientist*, 53(3):354–375. Publisher: SAGE Publications Inc.
- Li, H., Ajmani, L., Zhou, M., Vincent, N., Hwang, S., Piccardi, T., Narayan, S., Daniel, S., and Veselovsky, V. (2022a). Ethical Tensions, Norms, and Directions in the Extraction of Online Volunteer Work. In *Companion Publication of the 2022 Conference on Computer Supported Cooperative Work and Social Computing, CSCW’22 Companion*, pages 273–277, New York, NY, USA. Association for Computing Machinery.
- Li, H., Hecht, B., and Chancellor, S. (2022b). Measuring the monetary value of online volunteer work. *Proceedings of the International AAAI Conference on Web and Social Media*, 16:596–606.
- Lim, W. H., Carman, M. J., and Wong, S.-M. J. (2017). Estimating Relative User Expertise for Content Quality Prediction on Reddit. In *Proceedings of the 28th ACM Conference on Hypertext and Social Media*, HT ’17, pages 55–64, Prague, Czech Republic. Association for Computing Machinery.
- Lin, Z., Salehi, N., Yao, B., Chen, Y., and Bernstein, M. S. (2017). Better when it was smaller? Community content and behavior after massive growth. In *Eleventh International AAAI Conference on Web and Social Media*, pages 132–141, Montreal, Canada. AAAI.
- Lipset, S. M., Trow, M. A., and Coleman, J. S. (1956). *Union democracy: The Internal Politics of the International Typographical Union*. Free Press, Glencoe, Ill.
- Lo, C. C. W. Y. (2018). *When all you have is a banhammer : the social and communicative work of Volunteer moderators*. Thesis, Massachusetts Institute of Technology. Accepted: 2018-09-17T15:49:18Z.

- Long, K., Vines, J., Sutton, S., Brooker, P., Feltwell, T., Kirman, B., Barnett, J., and Lawson, S. (2017). "Could you define that in bot terms"?: Requesting, creating and using bots on reddit. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 3488–3500, New York, NY, USA. ACM. event-place: Denver, Colorado, USA.
- Ma, R. and Kou, Y. (2022). "I am not a YouTuber who can make whatever video I want. I have to keep appeasing algorithms": Bureaucracy of Creator Moderation on YouTube. In *Companion Publication of the 2022 Conference on Computer Supported Cooperative Work and Social Computing*, CSCW'22 Companion, pages 8–13, New York, NY, USA. Association for Computing Machinery.
- Ma, R., You, Y., Gui, X., and Kou, Y. (2023). How Do Users Experience Moderation?: A Systematic Literature Review. *Proc. ACM Hum.-Comput. Interact.*, 7(CSCW2):278:1–278:30.
- March, J., Schulz, M., and Zhou, X. (2000). *The dynamics of rules: change in written organizational codes*. Stanford University Press, Stanford, CA.
- Marvin, C. (1987). *When Old Technologies Were New*. Oxford University Press, Oxford, UK.
- Massanari, A. (2017). #Gamergate and The Fappening: How Reddit's algorithm, governance, and culture support toxic technocultures. *New Media & Society*, 19(3):329–346. tex.ids= massanari_gamergate_2017-1.
- Matias, J. N. (2016a). The civic labor of online moderators. *The Platform Society*, page 10.
- Matias, J. N. (2016b). Going dark: Social factors in collective action against platform operators in the Reddit blackout. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*, pages 1138–1151, New York, NY. ACM.
- Matias, J. N. (2019). The civic labor of volunteer moderators online. *Social Media + Society*, 5(2):1–12. tex.ids= matias_civic_2019-1.

- Meyer, J. W. and Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology*, 83(2):340–363.
- Monge, P. and Poole, M. S. (2008). The Evolution of Organizational Communication. *Journal of Communication*, 58(4):679–692.
- Morgan, J. T., Derthick, K., Ferro, T., Searle, E., Zachry, M., and Kriplean, T. (2009). Formalization and community investment in wikipedia’s regulating texts: the role of essays. In *Proceedings of the 27th ACM International Conference on Design of Communication*, SIGDOC ’09, pages 315–316, New York, NY, USA. ACM.
- Mugar, G. (2017). Preserving the Margins: Supporting Creativity and Resistance on Digital Participatory Platforms. *Proc. ACM Hum.-Comput. Interact.*, 1(CSCW):83:1–83:14.
- Mumby, D. K. and Stohl, C. (1996). Disciplining Organizational Communication Studies. *Management Communication Quarterly*, 10(1):50–72. Publisher: SAGE Publications Inc.
- Nagappa, A. (2023). Narratives of Change to Platform Governance on DTube, an Emerging Blockchain-Based Video-Sharing Platform. *Social Media + Society*, 9(3):20563051231196881. Publisher: SAGE Publications Ltd.
- Neff, G., Fiore-Silfvast, B., and Dossick, C. (2014). Materiality: Challenges and Opportunities for Communication Theory. SSRN Scholarly Paper ID 2435841, Social Science Research Network, Rochester, NY.
- Nicholson, M. N., Keegan, B. C., and Fiesler, C. (2023). Mastodon Rules: Characterizing Formal Rules on Popular Mastodon Instances. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing*, CSCW ’23 Companion, pages 86–90, New York, NY, USA. Association for Computing Machinery.
- North, D. C. (1990). *Institutions, institutional change, and economic performance*. Cambridge University Press, Cambridge, UK. tex.ids= north_institutions_1990-1.
- Orlikowski, W. J. (1992). The Duality of Technology: Rethinking the Concept of Technology in Organizations. *Organization Science*, 3(3):398–427.

- Orlikowski, W. J. (1996). Improvising organizational transformation over time: A situated change perspective. *Information Systems Research*, 7(1):63–92.
- Orlikowski, W. J. (2016). Sociomaterial Practices: Exploring Technology at Work. *Organization Studies*.
- Orlikowski, W. J. and Barley, S. R. (2001). Technology and institutions: what can research on information technology and research on organizations learn from each other? *MIS Q.*, 25(2):145–165.
- Orlikowski, W. J. and Gash, D. C. (1994). Technological frames: Making sense of information technology in organizations. *ACM Trans. Inf. Syst.*, 12(2):174–207.
- Orlikowski, W. J. and Scott, S. V. (2008). Sociomateriality: challenging the separation of technology, work and organization. *Academy of Management Annals*, 2(1):433–474.
- Osman, K. (2013). The role of conflict in determining consensus on quality in wikipedia articles. In *Proceedings of the 9th International Symposium on Open Collaboration, WikiSym '13*, pages 12:1–12:6, New York, NY, USA. ACM.
- Ostrom, E. (1990). *Governing the commons: the evolution of institutions for collective action*. Political Economy of Institutions and Decisions. Cambridge University Press, Cambridge.
- Pacey, A. (1983). *The Culture of Technology*. MIT Press. Google-Books-ID: JFfV7EopNPoC.
- Pentzold, C. (2011). Imagining the Wikipedia community: What do Wikipedia authors mean when they write about their ‘community’? *New Media & Society*, 13(5):704–721. Publisher: SAGE Publications.
- Pentzold, C. (2020). Social Norms and Rules in Peer Production. In *The Handbook of Peer Production*, pages 44–55. John Wiley & Sons, Ltd. Section: 4 _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781119537151.ch4>.

- Pfeffer, J. and Salancik, G. R. (1978). *The External Control of Organizations: A Resource Dependence Perspective*. Stanford University Press.
- Piskorski, M. J. and Gorbatai, A. D. (2017). Testing Coleman’s social-norm enforcement mechanism: Evidence from Wikipedia. *American Journal of Sociology*, 122(4):1183–1222.
- Posen, H. E., Keil, T., Kim, S., and Meissner, F. D. (2017). Renewing Research on Problemistic Search—A Review and Research Agenda. *Academy of Management Annals*, 12(1):208–251. Publisher: Academy of Management.
- Prasarnphanich, P. and Wagner, C. (2009). The Role of Wiki Technology and Altruism in Collaborative Knowledge Creation. *Journal of Computer Information Systems*, 49(4):33–41. eprint: <https://www.tandfonline.com/doi/pdf/10.1080/08874417.2009.11645338> tex.ids: prasarnphanich_role_2009 publisher: Taylor & Francis.
- Preece, J. (2000). *Online communities: designing usability, supporting sociability*. John Wiley, New York.
- Preece, J., Nonnecke, B., and Andrews, D. (2004). The top five reasons for lurking: Improving community experiences for everyone. *Computers in human behavior*, 20(2):201–223. tex.ids= preece_top_2004-1 place: Netherlands publisher: Elsevier Science.
- Putnam, L. and Mumby, D. K. (2014). *The Sage handbook of organizational communication: advances in theory, research, and methods*.
- Putnam, L. L. and Stohl, C. (1990). Bona fide groups: A reconceptualization of groups in context. *Communication Studies*, 41(3):248–265.
- Raman, A., Joglekar, S., Cristofaro, E. D., Sastry, N., and Tyson, G. (2019). Challenges in the decentralised web: the mastodon case. In *Proceedings of the Internet Measurement Conference, IMC ’19*, pages 217–229, New York, NY, USA. Association for Computing Machinery.
- Ransbotham, S. and Kane, G. C. (2011). Membership turnover and collaboration success in

- online communities: Explaining rises and falls from grace in Wikipedia. *MIS Quarterly*, 35(3):613–627.
- Reddy, H. and Chandrasekharan, E. (2023). Evolution of Rules in Reddit Communities. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '23 Companion, pages 278–282, New York, NY, USA. Association for Computing Machinery.
- Rheingold, H. (1993). *The virtual community: homesteading on the electronic frontier*. MIT Press, Cambridge, Mass, rev. ed edition.
- Ridings, C. M. and Gefen, D. (2004). Virtual Community Attraction: Why People Hang out Online. *Journal of Computer-Mediated Communication*, 10(JCMC10110).
- Ridings, C. M., Gefen, D., and Arinze, B. (2002). Some antecedents and effects of trust in virtual communities. *The Journal of Strategic Information Systems*, 11(3):271–295.
- Rijshouwer, E., Uitermark, J., and de Koster, W. (2023). Wikipedia: a self-organizing bureaucracy. *Information, Communication & Society*, 26(7):1285–1302. Publisher: Routledge _eprint: <https://doi.org/10.1080/1369118X.2021.1994633>.
- Ross, S. (2014). Your Day in 'Wiki-Court': ADR, Fairness, and Justice in Wikipedia's Global Community. SSRN Scholarly Paper ID 2495196, Social Science Research Network, Rochester, NY.
- Roth, C., Taraborelli, D., and Gilbert, N. (2008). Measuring wiki viability: An empirical assessment of the social dynamics of a large sample of wikis. In *Proceedings of the 4th International Symposium on Wikis*, WikiSym '08, pages 27:1–27:5, New York, NY. ACM.
- Rozenshtein, A. Z. (2023). Moderating the fediverse: content moderation on distributed social media. *Journal of Free Speech Law*, 3(1):217–236.
- Santos, F. M. and Eisenhardt, K. M. (2005). Organizational boundaries and theories of organization. *Organization Science*, 16(5):491–508.

- Schluger, C., Chang, J. P., Danescu-Niculescu-Mizil, C., and Levy, K. (2022). Proactive Moderation of Online Discussions: Existing Practices and the Potential for Algorithmic Support. *Proc. ACM Hum.-Comput. Interact.*, 6(CSCW2):370:1–370:27.
- Schneider, N. (2022). Admins, mods, and benevolent dictators for life: The implicit feudalism of online communities. *New Media & Society*, 24(9):1965–1985.
- Schneider, N., De Filippi, P., Frey, S., Tan, J. Z., and Zhang, A. X. (2021). Modular Politics: Toward a Governance Layer for Online Communities. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW1):16:1–16:26.
- Schoeneborn, D. and Vásquez, C. (2017a). Communicative Constitution of Organizations. In *The International Encyclopedia of Organizational Communication*, pages 1–21. American Cancer Society. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118955567.wbieoc030>.
- Schoeneborn, D. and Vásquez, C. (2017b). Communicative Constitution of Organizations. In *The International Encyclopedia of Organizational Communication*, pages 1–21. American Cancer Society. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118955567.wbieoc030>.
- Schöpke-Gonzalez, A. M., Atreja, S., Shin, H. N., Ahmed, N., and Hemphill, L. (2022). Why do volunteer content moderators quit? Burnout, conflict, and harmful behaviors. *New Media & Society*, page 14614448221138529. Publisher: SAGE Publications.
- Seering, J. (2019). Building More Positive Online Communities through Improving Moderation and Strengthening Social Identity. In *Companion Publication of the 2019 Conference on Computer Supported Cooperative Work and Social Computing, CSCW '19 Companion*, pages 89–93, New York, NY, USA. Association for Computing Machinery.
- Seering, J., Dym, B., Kaufman, G., and Bernstein, M. (2022a). Pride and Professionalization in Volunteer Moderation: Lessons for Effective Platform-User Collaboration. *Journal of Online Trust and Safety*, 1(2). Number: 2.

- Seering, J. and Kairam, S. R. (2022). Who Moderates on Twitch and What Do They Do? Quantifying Practices in Community Moderation on Twitch. *Proc. ACM Hum.-Comput. Interact.*, 7(GROUP):18:1–18:18.
- Seering, J., Kaufman, G., and Chancellor, S. (2022b). Metaphors in moderation. *New Media & Society*, 24(3):621–640. Publisher: SAGE Publications.
- Seering, J., Kraut, R., and Dabbish, L. (2017). Shaping Pro and Anti-Social Behavior on Twitch Through Moderation and Example-Setting. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, CSCW '17*, pages 111–125, New York, NY, USA. Association for Computing Machinery.
- Seering, J., Wang, T., Yoon, J., and Kaufman, G. (2019). Moderator engagement and community development in the age of algorithms. *New Media & Society*, 21(7):1–24.
- Shaw, A. and Hill, B. M. (2014). Laboratories of oligarchy? How the iron law extends to peer production. *Journal of Communication*, 64(2):215–238.
- Sherer, P. D. and Lee, K. (2002). Institutional Change in Large Law Firms: A Resource Dependency and Institutional Perspective. *The Academy of Management Journal*, 45(1):102–119.
- Singh, J. V., Tucker, D. J., and House, R. J. (1986). Organizational Legitimacy and the Liability of Newness. *Administrative Science Quarterly*, 31(2):171–193. Publisher: [Sage Publications, Inc., Johnson Graduate School of Management, Cornell University].
- Smirnov, I., Oprea, C., and Strohmaier, M. (2023). Toxic comments are associated with reduced activity of volunteer editors on Wikipedia. *PNAS Nexus*, 2(12):pgad385.
- Solomon, J. and Wash, R. (2012). Bootstrapping wikis: developing critical mass in a fledgling community by seeding content. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 261–264. ACM. Citation Key Alias: solomon_bootstrapping_2012.

- Srinivasan, K. B., Danescu-Niculescu-Mizil, C., Lee, L., and Tan, C. (2019). Content removal as a moderation strategy: compliance and other outcomes in the changemyview community. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):163:1–163:21.
- Stockinger, A., Schäfer, S., and Lecheler, S. (2023). Navigating the gray areas of content moderation: Professional moderators’ perspectives on uncivil user comments and the role of (AI-based) technological tools. *New Media & Society*, page 14614448231190901. Publisher: SAGE Publications.
- Suchman, M. C. (1995). Managing Legitimacy: Strategic and Institutional Approaches. *The Academy of Management Review*, 20(3):571–610. Publisher: Academy of Management.
- Tan, C. (2018). Tracing community genealogy: how new communities emerge from the old. In *Proceedings of the Twelfth International Conference on Web and Social Media (ICWSM ’18)*, pages 395–404, Palo Alto, California. AAAI.
- Tan, C. and Lee, L. (2015). All who wander: On the prevalence and characteristics of multi-community engagement. In *Proceedings of the 24th International Conference on World Wide Web, WWW ’15*, pages 1056–1066, Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee.
- TeBlunthuis, N. and Hill, B. M. (2022). Identifying competition and mutualism between online groups. In *International AAAI Conference on Web and Social Media (ICWSM 2022)*, volume 16, pages 993–1004, Atlanta, Georgia, USA. AAAI.
- TeBlunthuis, N., Kiene, C., Brown, I., Levi, L. A., McGinnis, N., and Hill, B. M. (2022). No community can do everything: why people participate in similar online communities. *Proceedings of the ACM on Human-Computer Interaction: Computer Supported Cooperative Work*, 6:1–25.
- TeBlunthuis, N., Shaw, A., and Hill, B. M. (2018). Revisiting ”The rise and decline” in a population of peer production projects. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 355:1–355:7, New York, NY. ACM.

- Tolbert, P. S. and Zucker, L. G. (1983). Institutional sources of change in the formal structure of organizations: the diffusion of civil service reform, 1880-1935. *Administrative Science Quarterly*, 28(1):22–39.
- Tosch, E., Garcia, L., Li, C., and Martens, C. (2024). Privacy Policies on the Fediverse: A Case Study of Mastodon Instances. *Proceedings on Privacy Enhancing Technologies*, 2024(4):700–733.
- Treem, J. W. and Leonardi, P. M. (2013). Social Media Use in Organizations: Exploring the Affordances of Visibility, Editability, Persistence, and Association. *Annals of the International Communication Association*, 36(1):143–189. Publisher: Routledge eprint: <https://doi.org/10.1080/23808985.2013.11679130>.
- Turco, C. J. (2016). *The conversational firm: Rethinking bureaucracy in the age of social media*. Columbia University Press, New York, NY.
- Uzzi, B. (1996). The sources and consequences of embeddedness for the economic performance of organizations: the network effect. *American Sociological Review*, 61(4):674–698.
- Van de Ven, A. H., Ganco, M., and Hinings, C. R. b. (2013). Returning to the Frontier of Contingency Theory of Organizational and Institutional Designs. *Academy of Management Annals*, 7(1):393–440. Publisher: Academy of Management.
- Van Maanen, J. (1978). The Asshole. In *Policing: A View from the Street*.
- Wadley, G., Smith, W., Ploderer, B., Pearce, J., Webber, S., Whooley, M., and Borland, R. (2014). What people talk about when they talk about quitting. In *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: the Future of Design*, OzCHI '14, pages 388–391, Sydney, New South Wales, Australia. Association for Computing Machinery.
- Walther, J. B. (1996). Computer-Mediated Communication: Impersonal, Interpersonal, and Hyperpersonal Interaction. *Communication Research*, 23(1):3–43.

- Wang, J. (2012). Survival factors for Free Open Source Software projects: A multi-stage perspective. *European Management Journal*, 30(4):352–371.
- Weber, M. (1978). *Economy and society*. University of California Press, Berkeley, CA.
- Weick, K. E., Sutcliffe, K. M., and Obstfeld, D. (2005). Organizing and the Process of Sensemaking. *Organization Science*, 16(4):409–421.
- Winner, L. (1980). Do Artifacts Have Politics? *Daedalus*, 109(1):121–136. Publisher: The MIT Press.
- Wohn, Y. (2019). Volunteer moderators in twitch micro communities: how they get involved, the roles they play, and the emotional labor they experience. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI’19, pages 1–13, Glasgow, UK. ACM.
- Yang, D., Kraut, R. E., Smith, T., Mayfield, E., and Jurafsky, D. (2019). Seekers, Providers, Welcomers, and Storytellers: Modeling Social Roles in Online Health Communities. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI ’19, pages 1–14, New York, NY, USA. Association for Computing Machinery.
- Yang, Y. (2019). When power goes wild online: How did a voluntary moderator’s abuse of power affect an online community? *Proceedings of the Association for Information Science and Technology*, 56(1):504–508. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/pra2.55>.
- Yao, M. Z. and Ling, R. (2020). “What Is Computer-Mediated Communication?”—An Introduction to the Special Issue. *Journal of Computer-Mediated Communication*, 25(1):4–8. Publisher: Oxford Academic.
- Yu, B., Seering, J., Spiel, K., and Watts, L. (2020). ”Taking Care of a Fruit Tree”: Nurturing as a Layer of Concern in Online Community Moderation. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI EA ’20, pages 1–9, New York, NY, USA. Association for Computing Machinery.

- Zhang, A. X., Hugh, G., and Bernstein, M. S. (2020). PolicyKit: Building governance in online communities. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, pages 365–378, Virtual Event USA. ACM. arXiv: 2008.04236.
- Zhu, H., Kraut, R. E., and Kittur, A. (2013). Effectiveness of Shared Leadership in Wikipedia. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 55(6):1021–1043.
- Zhu, H., Kraut, R. E., and Kittur, A. (2014). The impact of membership overlap on the survival of online communities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 281–290, New York, NY, USA. Association for Computing Machinery.
- Zulli, D., Liu, M., and Gehl, R. (2020). Rethinking the “social” in “social media”: Insights into topology, abstraction, and scale on the Mastodon social network. *New Media & Society*, 22(7):1188–1205. Publisher: SAGE Publications.

Appendix A

STUDY INFORMATION SHEET

Understanding Rules and Norms in Online Communities

You are invited to participate in a research study seeking to understand the reasons and processes by which online communities on different platforms adopt rules. Specifically, we seek to understand why communities often (but not always) post similar rules, how such rules shape the practices and experiences of online community leaders and participants, and how these processes are similar or different across several different platforms.

Eligibility to Participate

Participants must be adults (i.e., at least 18 years old if they are from the United States and/or above the age of majority in their locale if they are not), be able to conduct an interview in English, be familiar with online communities on at least one of the platforms we are studying, and be willing to participate in a 45-90 minute interview recorded over audio or video communication.

Procedures for the Study

If you agree to be in the study, you will first complete a short questionnaire asking about your familiarity and engagement with online communities. You will then participate in an interview conducted remotely by voice or video either using Zoom or a platform of your choice. This interview will include questions about your experience with [Reddit, Mastodon, Fandom, Open Source] communities.

You will be asked about your knowledge of these communities including their topics, content, written and unwritten rules, the kinds of people who are active in them and how they relate to one another. We expect that most interviews will last between 45 and 90 minutes. These interviews will be recorded and then transcribed so that our research team

can have a record of what you have said. We will delete the recordings themselves once the interviews are fully transcribed. We will edit out any identifiers from these transcripts so that all retained records are anonymous.

Voluntary Participation and Withdrawal

Taking part in this study is voluntary. You may choose not to take part or may leave the study at any time. For example, you may decline to complete the survey, cancel the interview before it begins or choose to end the interview at any point. Leaving the study will not result in any penalty or loss of benefits to which you are entitled. Your decision of whether or not to participate in the study will not affect your current or future relations with Northwestern University.

Payment for Participation

You may be compensated with a \$20 Amazon gift card. Amazon gift cards can only be used in the country where they were purchased and they can only be purchased on amazon.com's USA website or one of Amazon's international sites (Australia, Brazil, Canada, People's Republic of China, France, Germany, India, Italy, Japan, Mexico, Netherlands, Spain, United Kingdom). If you want to receive the gift card, we must collect your name, email address, and the country where you intend to use the gift card, in order to comply with financial requirements via this Google Form [link to form].

Confidentiality

We will endeavor to keep your personal information confidential. However, we cannot guarantee absolute confidentiality. Your personal information may be disclosed if required by law. Your identifying information including your email address, our records of which communities we asked you about, and audio/video recording of our conversation will be held in confidence until the reports resulting from the study are published, after which they will be deleted. Your username will be deleted after we successfully connect using email. The identities of communities you participate in will be removed from the transcripts.

Our university requires that if you opt to receive any payment gift card, we will need to collect your name and email address and retain this information for a period of 5-6 years. We will store this information in an encrypted file separate from the rest of the information we collect and will not share this information except in the case of a financial audit. If you do not wish to receive compensation, we will not collect your name or maintain any identifiers (like your email or username) after the study is concluded.

Questions or Problems

If you have questions, concerns, or complaints about the study or this information sheet, contact the researcher Aaron Shaw at aaronshaw@northwestern.edu

If you have questions about your rights as a research subject or if you think the research has hurt you, you may contact the Northwestern University Institutional Review Board Office at (312) 503-9338 or piper@northwestern.edu.

Appendix B

RECRUITMENT MESSAGES

Recruitment strategy: Reddit

To recruit eligible participants from Reddit communities, we will first identify Reddit communities with rules based on whether they publicly post rules through Reddit’s API and whether the rules are in English. We will download public rules pages for these eligible Reddit communities and sort them into groups based on similar rule text. Then, we will send recruitment messages to participate in our study to these Reddit communities’ Moderator Mail system. Eligible participants will have been an admin or moderator of the Reddit community and have been active at some point in time in the Reddit community’s moderation operations.

Recruitment message

The Community Data Science Collective (a research group based at Northwestern University and the University of Washington) is seeking individuals involved in a number of different online communities to participate in a research study on rules in online communities across platforms (Northwestern IRB # STU00217080). The research is led by PhD students Charles Kiene and Sohyeon Hwang and supervised by Principal Investigators Dr. Aaron Shaw (aaronshaw@northwestern.edu), Dr. Benjamin Mako Hill (makohill@uw.edu), and Dr. Nathan TeBlunthuis (nathan.teblunthuis@northwestern.edu).

We would love to schedule a time to talk with you about your experience with <Your Community> and other related online communities. The overall purpose of this study is to examine and understand how online communities like yours make their rules and how the rules shape the experience of participants.

In order for us to interview you, you must be an adult who is familiar with online communities on Reddit and who is willing and able to record an interview in English with

a researcher from our team. Interviews will be conducted remotely via voice or video chat depending on your preferences. While we may quote you in our study, we will not disclose any identifying information about you. We expect most interviews will be between 45 and 90 minutes.

If you are interested in taking part in this research study, please fill out this Google Form: [link to Reddit recruitment form](#).

Participants will receive a \$20 Amazon gift card.

More information on our research project is available on our website: <https://wiki.communitydata.science/>[Link to documentation](#).

Regards,

Sohyeon Hwang, Benjamin Mako Hill & Charles Kiene (/u/UWCOMResearcher)

Recruitment strategy: Fandom

To recruit members of Wiki communities hosted on the Fandom/Wikia platform, we will first identify communities that have sufficient rules and policies and then invite highly active community members to participate in our study. We will analyze publicly available data that the WikiTeam archivist group originally downloaded in compliance with Fandom's terms of use and that we have obtained from archive.org in order to identify potential communities and interviewees. We wish to recruit interviewees from communities representing a diversity of approaches to online community governance. Therefore, we will run a simple clustering analysis on the content of the policy pages and seek to recruit participants in communities from different clusters. We will recruit candidate participants by sending them direct messages on the Wiki platform using the following message:

Recruitment message

The Community Data Science Collective (a research group based at Northwestern University's Department of Communication Studies) is seeking individuals involved in a number of different online communities to participate in a research study titled "Rules in online communities across platforms" (Study ID# STU00217080). This research is being led by

Principal Investigator Dr. Aaron Shaw (aaronshaw@northwestern.edu) and Co-PI Nathan TeBlunthuis (myself; nathan.teblunthuis@northwestern.edu) at Northwestern.

We would love to schedule a time to talk with you about your experience with <Your Community> and other related online communities. The overall purpose of this study is to examine and understand how online communities like yours make their rules and how the rules shape the experience of participants.

In order for us to interview you, you must be an adult who is familiar with online communities on Fandom/Wikia and who is willing and able to record an interview in English with a researcher from our team. Interviews will be conducted remotely via voice or video chat depending on your preferences. While we may quote you in our study, we will not disclose any identifying information about you. We expect most interviews will be between 45 and 90 minutes.

If you are interested in taking part in this research study, please fill out this Google Form: <https://forms.gle/msqTRojgu3pi74W78>.

Participants who complete an interview will receive a USD \$20 Amazon gift card.

If you know anyone involved in creating policies on Fandom wiki who might be interested in participating in our study, please share this with them.

For more information on our research and who we are, visit our website: <https://wiki.communitydata.science/>.

Regards,

Nate TeBlunthuis

Recruitment

Recruitment strategy: Fediverse

We find active communities on the Fediverse based on a script we run daily to download the public metadata from public posts and to create a snapshot of community pages identified as important to description and rules. Using these public data, we first identify a set of communities which are primarily English-speaking and which show up in our sample as having some sort of rules or textual description. From these communities, we then sort

them into clusters based on textual similarity between the public metadata pages for each server, creating four main rules clusters. We identify administrators for these communities based on the public server metadata and moderators based on the about pages (if listed). We can also infer active members of the communities based on their activity if necessary using the public posts. Both the posts and metadata pages are public and do not require any login or special permissions to access.

Please note that because of post length restrictions on the Fediverse, we limit how much is in the message itself and link individuals to further relevant information (e.g. STU, PIs, etc.) on the linked wiki page. A copy of that page is uploaded to the eIRB+ system.

Recruitment language for initial message of recruitment:

Hello! I'm with the CDSC (a research group based at Northwestern University), seeking participants for an interview study (IRB STU#STU00217080) about rules in online communities.

We'd love to hear about your experience with <Your Community>.

If you are interested in taking part in this research study, please see https://wiki.communitydata.science/Fediverse_recruit.

Appendix C

INTERVIEW PROTOCOL

We will conduct a semi-structured interview lasting 45-90 minutes with each study participant generally following the structure below. We may ask participants to clarify their responses or to elaborate with greater detail through follow-up questions. Since this is a semi-structured interview that will rely on the flow of the conversation itself, we won't be able to anticipate all potential follow-up questions here. Similarly, we are unlikely to ask all questions listed below in any given interview. Some questions in the protocol are similar to one another and may appear redundant. This is intentional because interview participants may respond quite differently even to apparently similar questions.

We are conducting this study in several different settings consisting of different online platforms including Fediverse instances, Reddit.com, Fandom.com, and open source communities. Below, we refer to specific communities and platforms using generic placeholders such as <Community A> or <Platform B>. In the actual interviews we will substitute these placeholders for the actual relevant community name. Some questions address specific affordances of platforms and as a result may not make sense in other contexts.

Consent

Who is the interviewer, what are they doing?

1. My name is <name> and I'm from the [University]. Thank you for spending the time to talk with me today. I would like to talk to you about your experiences with rules and moderation in online communities. You should have already received an email with a Study Information Sheet containing detailed information about this study.

Privacy and confidentiality

To clarify, your participation in this conversation is strictly confidential—your responses may be published as quotes in a future academic research paper, but no identifiers will tie your username or any other identifiable factors to those responses. There are no wrong answers to my questions. I want to hear your personal stories and your experiences. You also absolutely can decline to answer any of the questions that I ask. You are also free to stop the interview at any time. This is your interview.

Participant questions

Do you have any questions about this interview before we start?

Consent to interview and recording

- For the record, is it okay if I record this conversation?
- (If they say yes, start recording)
- (Explanation of the open interview structure)

Questions for interview participants

Questions about participation

1. What brought you to community <Community A>?
2. How did you choose to participate in <Community A>?
3. How long have you been involved with <Community A>?
4. How do you see your role within <Community A>?
5. How would you describe <Community A>?
6. Are you paid or otherwise compensated for your participation in <Community A>?

7. Do <Community A>'s choice of rules influence your decisions to choose to join or participate in <Community A> in any way? How so?
8. Have you ever had a negative experience with <Community A>? If so, was it related to the rules in any way?

Questions about rules in general

1. Can you tell me about <Community A>'s rules?
2. What do you believe is the purpose of rules in <Community A>?
3. Can you confirm that <Community A> uses the rules on their website?
4. Were you aware of <Community A>'s rules before you signed up?
5. Can you recall some specific rules of <Community A>? Why do these ones stick out to you?
6. How important are the rules in <Community A>?
7. What do you think would happen if you didn't have <rule X>?
8. Are there any unwritten rules or norms in <Community A>?
9. Are there any rules that <Community A> has that you wish it didn't? If so, why do you think they still have this rule?
10. How do newcomers learn the rules of <Community A>? How did you learn of the rules of <Community A> when you first joined?
11. Can you walk me through the first time you interacted with <Community A>'s rules / how you became familiar with <Community A>'s rules?
12. What kinds of behaviors does <Community A> view negatively? Why?

13. What kinds of behaviors does <Community A> view positively? Why?
14. (In follow up to the two questions above) How do <Community A>'s rules address <Behavior Y>?
15. Are there any rules that you wish that <Community A> had? Why? If so, why do you think the community doesn't have a rule like that?
16. How do <Community A>'s rules align with your personal values? How do these rules reflect your values or not?
17. What are your personal values in community participation?
18. Can you tell me about your personal approaches to community participation?
19. How do <Community A>'s rules shape your contributions to this community? For example, do the rules affect the types of content or work that you produce and/or share?
20. In what types of settings or activities are your rules applied? Do the same rules apply in the mailing list as in the chat room? Do the same rules apply if you meet up in-person or at a convention?
21. Do you have different rules for different mediums or contexts? (e.g., mailing lists, chat rooms, forums, in-person meetups)
22. Are there things that tend to be agreed upon within <Community A> that aren't written in formal rules?
23. How do the rules help, hurt, or otherwise impact the community?
24. Can you describe any problems in <Community A>?
25. Do the rules attempt to deal with these problems in any way?

Questions about the origins of rules

1. How did <Community A> choose its rules?
2. Why did you choose these rules?
3. Have you ever had to change the rules?
4. How have the rules changed, if ever?
5. Are there any other communities that <Community A> looked to when you were choosing rules?
6. What is the most recent rule that <Community A> has adopted? Tell me about that?
7. How do you think <Community A> decided to phrase <Rule X>?
8. Are there any rules that <Community A> used to have, but doesn't have anymore? If so, do you know why this change was made?
9. What is the process for adopting or changing rules in <Community A>?
10. Do you know who set the rules in <Community A>? Can you tell me more about that?
11. Where do [you think] these rules come from? (looking at/through the rule list together)

Questions for moderators, administrators, or other leadership positions

1. Is moderating <Community A> part of your job or otherwise a paid activity?
2. Why do you work as a moderator of <Community A>?

3. How did you become a moderator of <Community A>?
4. How would you describe your work as a moderator of <Community A>?

Questions about the enforcement of rules

1. What do you think <rule X> means in practice? How does that feel different to what the rule means in theory, if at all?
2. How many people are involved in rule enforcement? What roles do they hold within the community?
3. Do you feel as though the rules are fairly/equally enforced in <Community A>? (Follow up: why do you think that?)
4. Do you feel as though the rules are consistently enforced in <Community A>? (Follow up: why do you think that?)
5. Do you feel as though it is possible for rules to be effectively enforced in <Community A>? (Follow up: why do you think that?)
6. How do you think moderators/administrators of <Community A> use the rules?
7. How strictly do the moderators of <Community A> enforce the rules? Why? What makes you feel like it is strict or not?
8. Do ordinary members of the community help enforce the rules or norms? (e.g., "call each other out")? If so, can you share an example of a time that happened?
9. If not, why do you think not?
10. In what ways are rules or norms enforced in <Community A>? Can you share an example of a time that happened?

11. Are there particular people who take responsibility for enforcing the rules or norms?
Are these people normally in formal roles, like moderators or administrators?
12. What happens when there is conflict over how rules are enforced (internal to rules enforcement team)?
13. What happens when there is conflict over how rules are enforced (external to rules enforcement team/community wide)?
14. Are there any rules for moderators specifically?

Questions about related communities

1. Are you a part of other online communities?
2. How do the rules of <Community A> compare to the rules of related communities?
3. Do all communities like yours have a rule like <rule X>?
4. Can you name some other communities with rules?
5. Can you compare the rules of <Community A> to <Community B>?
6. How do the problems of <Community A> compare to those of <Community B>?
7. Do you think differences in rules shape the relationship between <Community A> and <Community B> in any way?

Questions about rules and platforms

1. Do you feel like you need to have certain rules because you are on <Platform A>?
2. How important are the rules on <Platform A>?

3. Can you tell me about the [most important] rules on <Platform A>? How does that translate into <Community A>?
4. How do <Community A>'s rules align or not with <Platform A>'s rules?
5. How do you think about <Platform A>?
6. How does <Platform A> compare with other online community platforms?
7. Are there other communities on <Platform A> you engage with regularly?
8. Do you feel as though the platform rules protect you personally?
9. Who do your community rules protect?

Questions about Background Demographics

1. Do you identify as being a minority or otherwise marginalized within your culture?
2. Do you identify as a minority or otherwise marginalized within <Community A>?
3. How does your identity or background shape your experience with <Community A>?
4. Do you feel like others who share your identity or background would feel comfortable participating in <Community A>? Why or why not?
5. Would you recommend <Community A> to people with minority or otherwise marginalized identities? Why or why not?
6. Can you describe any ways that the rules or norms of <Community A> could do more to help those with your identity or background?
7. Can you describe any ways that the rules or norms of <Community A> already help those with your identity or background?

8. Have you ever had a negative experience with <Community A> related to your identity or background?

General follow-up questions

1. Can you say more about that?
2. Could you explain what you mean by ...?
3. What made you feel [a certain way]?
4. Is there an incident you're thinking of? Can you recall an incident or example for me?
5. Can you tell me what led to (the event)?
6. What was your role in this event?
7. What happened after?
8. What were your thoughts then? What are your thoughts now?
9. How so? What about (alternative)?
10. Could you tell me about one of those times? (assuming they say something general, and we want a specific example)
11. Could you go into more detail about (this incident, this comment, why you think this)?
12. Have you found that this is common in your online experiences? Does anything strike to be different about this experience in this community?
13. On a scale of one to ten, how important is this to you?

Appendix D

PARTICIPANT SCREENING SURVEY

Research Study Information

Title of Research Study: Rules in Online Communities Across Platforms (IRB STU#STU00217080)

Principal Investigator: Aaron Shaw

Supported By: This research is supported by the National Science Foundation.

Collaborating Institutions: University of Washington (Principal Investigator: Benjamin Mako Hill)

Instructions:

Please answer the following questions. Fields marked with an asterisk () are required.*

1. Username*

Please enter the username by which we contacted you.

2. Email Address (Fandom only)

Please provide an email address where we may contact you.

3. Are you an active participant in the online community where we contacted you?*

- Yes, I am an active participant.
- No, I am not.

4. Do you have knowledge of the community's rules, norms, and governance?*

- Yes, I am familiar with the community's rules, norms, and governance.

- No, I am not.

5. *Are you an adult (above the age of majority in your country or locale)?**

- Yes, I am an adult.
- No, I am a minor.

6. *Do you speak English?**

- Yes, I speak English.
- No, I do not.

7. *Are you willing and able to complete a 45-90 minute virtual interview in English?**

- Yes, I am willing and able to complete an interview in English.
- No, I am not.

8. *Do you understand that compensation for participation in this research is available in the form of Amazon Gift Cards (\$20 USD) and that these gift cards will be usable only in one of the following countries: USA, Australia, Brazil, Canada, People's Republic of China, France, Germany, India, Italy, Japan, Mexico, Netherlands, Spain, United Kingdom?**

- I understand.
- I do not understand.

9. *What is your current gender?**

Please select or specify your gender.

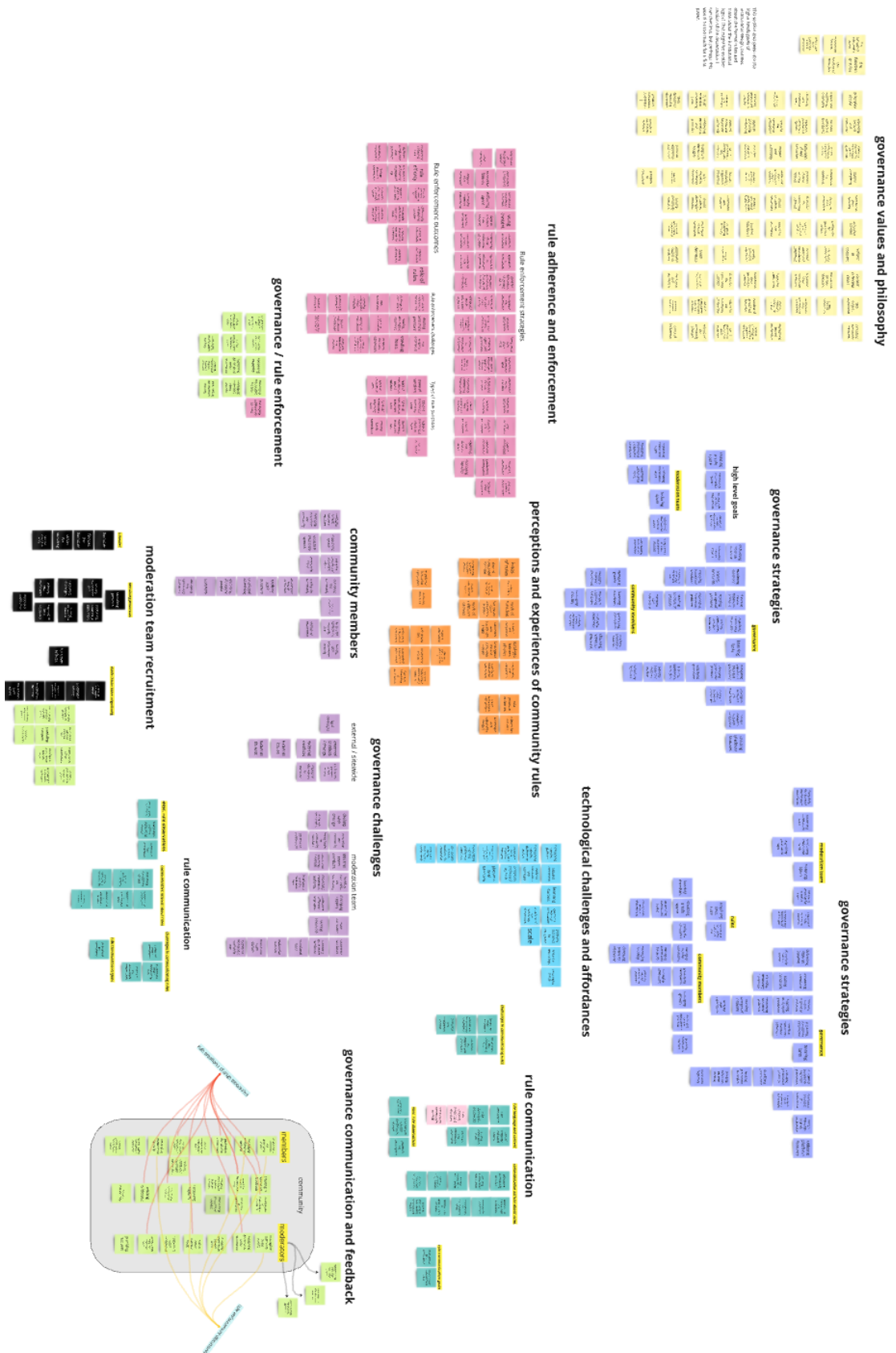
10. *What is your educational background?**

- I did not complete high school or secondary education.
- I completed high school or secondary education.
- I have some university or college education.
- I completed university/college.
- I have an advanced degree, like a medical, legal, or other professional degree, a master's degree, or a PhD.

Thank you for participating!

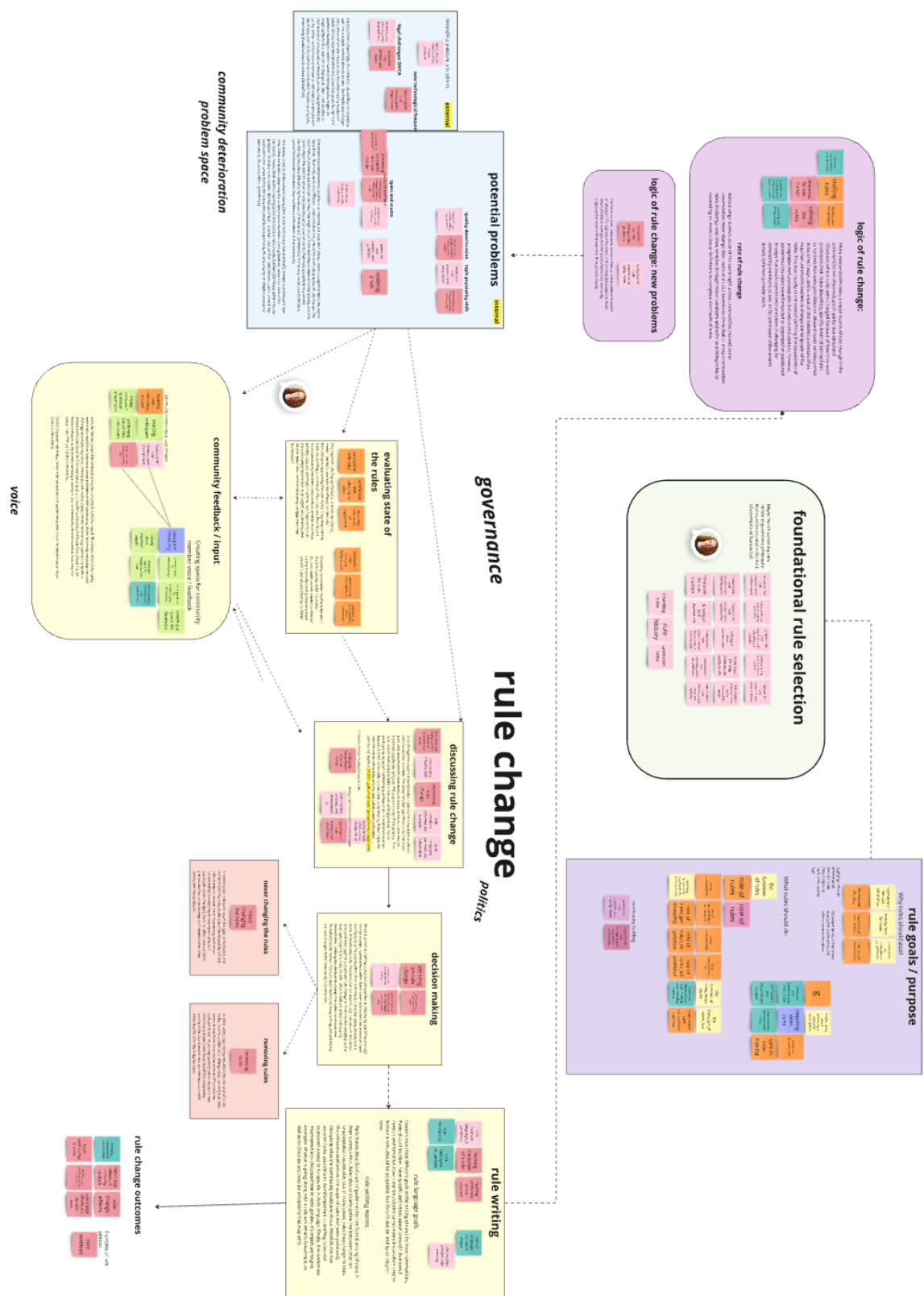
Appendix E

FOCUSED CODE ORGANIZATION



Appendix F

SECOND FOCUSED CODE ORGANIZATION



Appendix G

RULE CHANGE PROCESS

