The Civic Labor of Volunteer Moderators Online

J. Nathan Matias

Abstract
Volunteer moderators create, support, and control public discourse for millions of people online, even as moderators’ uncompensated labor upholds platform funding models. What is the meaning of this work and who is it for? In this article, I examine the meanings of volunteer moderation on the social news platform reddit. Scholarship on volunteer moderation has viewed this work separately as digital labor for platforms, civic participation in communities, or oligarchy among other moderators. In mixed-methods research sampled from over 52,000 subreddit communities and in over a dozen interviews, I show how moderators adopt all of these frames as they develop and re-develop everyday meanings of moderation—facing the platform, their communities, and other moderators alike. I also show how this civic notion of digital labor brings clarity to a strike by moderators in July 2015. Volunteer governance remains a common approach to managing social relations, conflict, and civil liberties online. Our ability to see how communities negotiate the meaning of moderation will shape our capacity to address digital governance as a society.

Keywords
online behavior, digital labor, Internet governance, collective action, content moderation

Introduction
On 2 July 2015, volunteer moderators of over 2,200 “subreddit” communities on the social news platform reddit effectively went on strike. Moderators disabled their subreddits, preventing millions of subscribers from accessing basic parts of the reddit website. The “reddit blackout,” as it became known, choked the company from advertising revenue and forced reddit to negotiate over moderators’ digital working conditions. The company, already struggling with pressure from racist and bullying groups that it had recently banned, conceded to moderator demands within hours. Management allocated resources to moderator needs, CEO Ellen Pao resigned 1 week later, and within 2 months, the company had hired its first Chief Technical Officer, partly to improve the platform’s moderation software (Olanoff, 2015).

Even as the blackout surfaced anxieties about the responsibilities of digital platforms to their volunteer workers, it also led many to question the legitimacy of moderators’ governance role. Some moderators were censured or even ejected by their subreddits for joining the blackout without consulting their communities. Conversely, many moderators were pressured to join the blackout through subreddit-wide votes and waves of private messages. Three weeks later, in The New York Times magazine article, on the word “moderator,” Adrian Chen (2015) wrote,

The moderator class has become so detached from its mediating role at Reddit that it no longer functions as a means of creating a harmonious community, let alone a profitable business. It has become an end in itself—a sort of moderatocracy.

Are these moderators unpaid workers whose emotional labor is exploited by platforms, are they facilitator citizens upholding society’s collective communications, or are they oligarchs who coordinate to rule our online lives with limited accountability? Chen struggles to reconcile these views for good reason. When making sense of the work of moderation, scholars have tended to think primarily in one of three ways. Scholarship on digital labor describes moderation as unwaged labor for commercial interests or free labor in peer production communities like Wikipedia (Menking & Erickson, 2015; Postigo, 2003; Terranova, 2000). Legal theorists and computer scientists describe moderators as civic leaders of online communities who build their own public spheres (Kelty, 2005); much of this scholarship outlines...
general strategies to structure governance work for fair and functional communities at scale (Butler, Sproull, Kiesler, & Kraut, 2002; Grimmelmann, 2015). A third conversation draws from the sociology of participation to consider the social structures of those who acquire and exercise moderation power, finding common tendencies toward oligarchy that may be necessary for the survival of online communities (Shaw & Hill, 2014; Zhu, Kraut, & Kittur, 2014).

Even as scholars debate the nature of moderation work, online communities routinely define what it means to be a moderator in everyday settings: they dispute over moderator decisions, recruit new moderators, participate in elections, investigate corruption, offer mentorship, and share peer support. In their everyday work, moderators must satisfy and explain themselves to all three parties identified in previous research, sometimes simultaneously: the platform, their communities, and their fellow moderators. The platform operators must be satisfied that a moderator is appropriately productive, communities must accept the legitimacy of a moderator’s governance, and other moderators must also trust and support the moderator throughout their work.

Academic views of moderation work typically attend to only one of these stakeholders at a time. Digital labor research on the role of moderation in a “profitable business” attends to the relationship between moderation work and platform operators. Scholarship on the civic outcomes of moderation emphasizes the relationship of moderators with the publics they govern. Finally, studies on moderator social structures draw attention to the ties and obligations of moderators to each other.

The everyday work of defining volunteer moderation is central to the legitimacy and power of online governance; however, scholars choose to describe it. Consider, for example, the issue of compensation. Since moderators create and enact policy on acceptable speech, their work fundamentally shapes our digitally mediated social and political lives. Moderators respond to conflict and harassment online, risks that 40% of American adults report experiencing (Duggan, 2014). This valuable work is costly. Professional services reportedly charged between US$4 and US$25 cents per comment in 2014 (Isaf, 2014). In 2008, America Online (AOL) community leaders settled a class action lawsuit over unpaid wages for US$15 million (Kirchner, 2011). In recent years, many news organizations have disabled public discussions, unable to afford moderation costs (Gupta, 2016).

Although platforms could afford moderation costs, the legitimacy of moderation is also affected by how communities interpret compensation models. On reddit, many communities see paid moderation as corruption, forcing out moderators accused of receiving compensation or favors in exchange for their labor (Martinez, 2013). Because moderation is governance as well as labor, its legitimacy depends on the beliefs of people other than the moderators who create and enforce policies. Consequently, the processes that shape the meaning of moderation also define its power.

In this article, I examine how the meaning of moderation is defined in the everyday boundary work carried out by volunteer moderators on reddit as they negotiate the idea of moderation. Boundary work, as described by Gieryn, is discursive activity that attempts to define the boundaries of a profession or field, to support claims to authority and resources (Gieryn, 1983). These boundaries are “drawn and redrawn in flexible, historically changing and sometimes ambiguous ways” that reflect the ambivalences and strains within a given institution. In online platforms such as reddit, volunteer moderators define and redefine what it means to be a moderator in conversation with platform operators, their communities, and other moderators. To foreground the ways that moderation is defined with all three parties, I introduce the idea of “civic labor” to describe authority that is defined through negotiations with these commercial, civic, and peer stakeholders.

**Moderation Work**

While online platforms do pay some people to enact their content policies (Gillespie, 2018; Roberts, 2016), volunteer moderators have played a fundamental role in social life online for over 40 years. Many online social systems fundamentally rely on volunteers, from librarians in 1970s Berkeley looking after local message-boards (Bruckman, 1998) to today’s Facebook group administrators (Kushin & Kitchener, 2009), Wikipedia arbitrators (Menking & Erickson, 2015), and reddit moderators. Although not all work of fostering community is carried out by designated moderators, people in these formal positions are founders, maintainers, content producers, promoters, policymakers, and enforcers of policy across the social Internet (Butler et al., 2002). On many platforms, moderators also manage autonomous and semi-autonomous moderation software that work alongside them (Geiger & Ribes, 2010).

By delegating policy and governance power to moderators, platform operators reduce labor costs and limit their regulatory liability for conduct on their service while also positioning themselves as champions of free expression and cultural generativity (Gillespie, 2010). This governance work invites public scrutiny, which draws platforms into debates about their responses to flagged material (Crawford & Gillespie, 2014). However, when platforms delegate policy-making to their users, that scrutiny is faced instead by moderators, whose labor nonetheless upholds a platform’s economic model.

On reddit, the evolution of moderation followed this longer 40-year pattern. When reddit’s creators founded it in 2005 to be “the front page of the Internet,” they developed an infrastructure for sharing and promoting highly voted posts a single, algorithmically curated page. After these algorithms regularly promoted pornography and other complicated, possibly illegal material, the platform created an alternative algorithmic space for “Not Safe For Work” (NSFW) material,
calling it a “subreddit” 1 month later (Huffman, 2006). Over the next 2 years, the company started dozens of new subreddits, mostly to separate conversations in different languages. In January 2008, after its acquisition by Condé Nast and 10 months after introducing advertising, the company launched “user-controlled subreddits.” Before that, users could join official company subreddits, reporting spam and abuse directly to the company through a flagging system. Now they could create their own public and private subreddits, taking action themselves to “remove posts and ban users” (Huffman, 2007, 2008). By giving communities delegated power to define their own governance, reddit was positioning itself as a platform and disclaiming responsibility for how its users behaved.

Seven years later, reddit was one of the largest social platforms online. In the month before the reddit blackout, the company received over 160 million visitors, roughly half of the number of active Twitter users in the same period. To maintain social relations at that scale, reddit relied on nearly 150,000 moderator roles for over 52,000 monthly active subreddits.

**Moderation as Free Labor in the Social Factory of Internet Platforms**

Digital labor scholarship on the work of moderators foregrounds their relationship with online platforms: theorizing the role of moderators' volunteer work within platform business models. Among examples in open source and free culture, this scholarship also frequently refers to labor organizing by community leaders (essentially moderators) of AOL chat rooms and other communities in the 1990s. Initially eager to offer moderation work in exchange for discounts, credit, and other perks, some of the 14,000 “community leaders” came to see their work as unpaid labor. Moderators filed a class action lawsuit in 1999, prompting an inconclusive US Department of Labor investigation. The community leaders eventually won US$15 million from AOL in a 2008 settlement (Kirchner, 2011; Postigo, 2009).

In an analysis of labor organizing by AOL moderators, Terranova points out that this freely given labor comprises an arrangement where people carry out self-directed cultural and social work that produces the value extracted by platforms. For Terranova, the “free labor” of platform production is something that is both “not financially rewarded [by platforms] and willingly given [by users]” (Terranova, 2000).

In a series of articles on the AOL lawsuit, Postigo explores the nature of the delicate symbiosis between platforms and moderators by observing the factors that led this arrangement to collapse. Postigo observes that the gift of volunteer time by AOL moderators was inspired by the “early Internet community spirit” found in “hacker history” and in “the academic, collaborative efforts that shaped the Internet” in the 1960s, 1970s, and 1980s. Yet some also took on the role to grow their technical skills or gain the discounts initially offered to volunteers. As AOL grew, the company began to formalize and control the relationship with their community leaders through communications, software, and compensation structures. No longer allowed the autonomy to imagine themselves as cultural gift-givers, the community leaders reimagined themselves as mistreated employees and sued the company. Postigo describes their labor organizing as an effort to “stake out new occupational territory” for “community making” on the Internet, an example of people who were “breaking out of the ‘social factory’” that Terranova put forward (Postigo, 2003, 2009).

Terranova and Postigo rightly draw attention to the co-dependence of many online platforms with the substantial uncompensated labor that continues to support them. Community management is now more common as a paid position, but the majority of the labor continues to be unpaid. Theories of digital labor offer clarity on the challenges of creating a “profitable business,” through volunteer labor, as Adrian Chen phrased it in The New York Times.

In many ways, the reddit blackout defies explanation by prior theories of volunteer moderation. Moderators did not attempt to stake out their work as an occupation, nor did they demand compensation. Instead, they leveraged reddit’s dependence on advertising to force the company to better meet their needs and those of their communities. As Centivany has argued, the reddit blackout was a social movement focused on company policy, a moment where the dependence of a platform on volunteer labor was deployed to achieve aims with as many civic dimensions as economic ones (Centivany & Glushko, 2016).

**Moderation as Civic Participation**

Volunteer moderation is also the work of creating, maintaining, and defining “networked publics,” imagined collective spaces that “allow people to gather for social, cultural, and civic purposes” (boyd, 2010). While social platforms offer technical infrastructures that constitute these publics, the work of creating and maintaining these imagined spaces is carried out in many everyday ways by platform participants and moderators. Butler and colleagues call the work of moderation “community maintenance,” drawing attention to the “communal challenge of developing and maintaining their existence.” They compare these communities to neighborhood societies, churches, and social movements. Writing about the details of community work online, Butler and colleagues draw attention to the benefits of affiliation and social capital. Where Terranova and Postigo see labor in service of platform business models, Butler and colleagues (2002) describe community maintenance as a service to the community itself. This view on the work of maintaining communities is similar to what Boyte and Kari (1996) call “public work,” an activity of cooperative citizenship that “creates social as well as material culture” (p. 21). Aside from the unique challenges of tending community software,
the mailing list moderators studied by Butler support their communities by recruiting newcomers, managing social dynamics, and participating in the community.

As online harassment has grown in prominence, scholarship on the role of moderators has drawn attention to their work to protect people’s capacities to participate in publics. Volunteers who respond to harassment create and manage technical infrastructures such as “block bots” and moderation bots to filter “harassment, incivility, hate speech, trolling, and other related phenomena,” argues Stuart Geiger. These volunteer efforts see moderation as “a civil rights issue of governance,” where marginalized groups deploy community infrastructure to claim spaces for conversation, community, and support (Geiger, 2016).

While these civic perspectives on moderation acknowledge the role of platforms, they foreground the relationship between moderators and the publics they are responsible for. The labor of moderators does sustain platform economies, yet the work itself is most directly concerned with the specific communities they govern. When moderators are questioned, as Adrian Chen did in The New York Times magazine, it is often for their record at fostering “harmonious community.” Yet theories of moderation as civic participation miss important ways that moderators define their work in relation to platforms and other moderators, sometimes in ways that conflict with the wishes of their communities.

**Moderation as Oligarchy**

Even as moderation work supports community, the power of individual moderators is defined and managed by other moderators who gate-keep the process of taking on and maintaining the role. A third perspective on volunteer moderation examines ways that this work is socially structured by other moderators and the interests of these moderators can diverge from the goals of their communities.

Early theories of leadership development in online communities imagined a “reader to leader” process where more active participants gain greater responsibility over time (Preece & Shneiderman, 2009). However, longitudinal research by Shaw and Hill has shown online communities to be much more like other voluntary organizations, where “group of early members consolidate and exercise a monopoly of power within the organization as their interests diverge from the collective’s.” Across 683 Wikia wikis, they find support for this “iron law of oligarchy,” showing that on average, a small group does come to control the positions of formal authority as a wiki grows (Shaw & Hill, 2014). Yet where Shaw and Hill see oligarchy, others see experience necessary for online communities to flourish. Also studying Wikia, Zhu and colleagues (2014) interpreted similar findings to argue that communities whose leaders also lead other communities are more likely to survive and grow.

In all these cases, experienced and powerful moderators control the process for others to gain and maintain their positions. Anyone seeking the role must negotiate that position with other moderators as well as their community and the platform. While moderators are powerful as a group, theories of oligarchy cannot explain the ways that platforms and communities do exert power in volunteer moderation, or the ways that moderators negotiate their work in relation to those other stakeholders.

**Standpoint and Methods**

My attempt to understand the meaning of volunteer moderation is grounded in my standpoint as a researcher who works directly with online communities and volunteer moderators in studies that are independent from the technology industry (Matias & Mou, 2018). When developing this research, I needed ways to think about the power relations of volunteer moderation and how to negotiate that power with the stakeholders involved. I began asking these questions after leading a team to study efforts by Women, Action, and the Media (WAM!), a non-governmental organization (NGO) that was supporting people experiencing harassment on Twitter (Matias et al., 2015). The volunteers who reviewed harassment reports and advocated the cases to Twitter were criticized from multiple directions. Some argued that these advocates represented a step backward for progress on online harassment, taking on labor that Twitter should be paying for (Meyer, 2014). WAM! certainly managed its relationship with Twitter to retain the privilege of supporting harassment receivers and maintain a public voice on the company’s policies. Others called our project a dangerous form of authoritarian censorship (Sullivan, 2014). The volunteers saw their work as a contribution to civic life in service to the people who asked for their help. Which of these was true? In our answers to ourselves and to these stakeholders, WAM! and our research team needed to draw and redraw the boundaries of our work to manage public expectations and serve the public good we hoped we could provide.

My fieldwork with reddit moderators began at a time when I was trying understand the many-sided scrutiny that WAM!’s harassment reviewers had faced. WAM!’s respondents might be unpaid volunteers who took on a substantial burden of emotional labor, but they were also a privately selected group with substantial power over others. Their work served platform operators who could remove them at will. They also served and governed users, who pressured them to share and justify their actions. As I spent time with reddit moderators, I watched them respond to similar questions from these multiple sides, a position many moderators had been negotiating for years.

To study the discursive boundary work that reddit moderators conduct with platforms, communities, and each other, I carried out participant observation, content analysis, interviews, and trace data collection on the social news site reddit over a 4-month period from June through September 2015, with follow-up data collection through February 2016.
Collected content includes 10 years of public statements by the company, 90 published interviews by moderators of other moderators, statements by over 200 subreddits that joined the blackout, over 150 subreddit discussions after concluding participation in the blackout, and over 100 discussions in subreddits that declined to join the blackout. I also used the reddit API to conduct trace analysis of moderator roles in the population of 52,735 active subreddits. Finally, I held semi-structured interviews with 14 moderators of subreddits of all sizes, sampled from communities both sides of the blackout. Interviewees included moderators of “NSFW” subreddits only available to users 18 years or older, as well as more widely accessible subreddits. Moderators of subreddits allegedly associated with hate speech declined to participate. I coded interviews, blog posts, online discussions, and other records by entering them into the Tinderbox information management system, where I tagged, clustered, and constructed qualitative evidence (Bernstein, 2003).

In this article, I focus on moments of tension and transition that brought debates over the meaning of moderation to the fore, including disputes over moderator decisions, the process of becoming a moderator, transitions of leadership, conflicts between communities, crises of legitimacy, the work of starting new communities, debates over compensation, and collective action during the reddit blackout of July 2015. Throughout points of tension and transition, moderators carry out the work of defining this civic labor at the boundaries of their relationships with platforms, their communities, and other moderators.

Disputing and Justifying Moderation Decisions with Communities

When someone’s contribution to reddit is removed by moderators, it can often come as a surprise. Since many participants engage primarily with the platform’s aggregated feed, they may not be aware that the posts they submit are subject to a subreddit’s community policies (Massanari, 2015). Responses to moderation decisions are often received through “modmail,” a shared inbox for each subreddit’s moderators. Complaints often include moderation policy debates, profanity, racist slurs, and threats of violence. Even when moderators ignore the complaints, these disputes shape the language the moderators use to describe their roles as dictators, martyrs, janitors, hosts, connoisseurs, and policymakers.

Some moderators describe themselves as “dictators,” arguing that the power they exercised needed no justification. In these communities, “the top mod makes all the decisions, usually because s/he created the sub.” Those who complain are urged either to accept moderator power or to stay away.

Moderators of subreddits dedicated to marginalized communities sometimes explain themselves as defenders. One moderator described the former moderator of a gender minority subreddit as a “martyr, angry and whirling and ready to give hell to anyone who dared to cross her or to threaten her communities.” When adopting the figure of a defender, moderators draw attention to the moral and political justifications for their exercise of power.

Other moderators adopt language from hospitality or service labor, describing themselves as “hosts” and “janitors.” These analogies de-politicize their role. Describing themselves in this way, one moderator argued that “my subreddits belong to my communities, I just happen to help out by cleaning up.” Reflecting on the accusations and complaints they receive, another moderator explained,

> It seems like it’s some sort of important position, while it’s actually just janitorial work . . . the degree of accusations, insults, abuse and unreasonable complaints from the politically interested is extreme . . . it’s janitorial when you remove hundreds of comments that just say “kill yourself blackie.”

When I asked moderators whether the language of janitor also implied a labor critique toward the reddit company, they disagreed. One described the language of janitor as “a response to complaints about conspiracies, censorship, etc” rather than their relationship to the company.

Many moderators describe themselves as connoisseurs when explaining their decisions about what to remove. In one subreddit dedicated to shocking material, moderators expressed disappointment over the lack of nuance and quality in submitters’ sense of the truly shocking. For example, one moderator claimed that too many submitters are shocked by images of nudity, violent injury, or death; moderators considered these too commonplace for inclusion. These moderators described themselves as taste-makers for their communities: “we are fucked up, but in a courtesy sniff kinda way that you’re ok with sharing with your friends.”

Some moderators respond to complaints of censorship by drawing inspiration from the language of governance. These subreddits describe their decisions in terms of “policies” and sometimes produce transparency reports of moderation actions. One subreddit described its transparency report as a response to participant complaints, an effort “towards improving user-moderator relations.” Their five-page report offered an empirical response to common complaints received by moderators of this 10 million subscriber community. Several other large subreddits publish aggregated transparency reports, with some sharing public logs of every action taken by the group’s moderators. By publishing transparency reports, moderators position themselves as civic actors accountable to their communities. The reports deflect criticism while also inviting evidence-based discussions of moderation practices.

The language of governance is also used by reddit participants who investigate and analyze moderator behavior. One interviewee described investigating and “exposing” a
Moderator for encouraging reddit users to share sexual photographs of minors. The investigators organized a press campaign to pressure the company, who then shut down the subreddit involved (Morris, 2011). In another case, participants accused a large technology subreddit’s moderators of censoring political discussions. To support these accusations, one person conducted data analysis of the subreddit’s history, creating charts that showed a sharp cutoff in discussions of surveillance and other political topics. The moderators’ accusers argued that the subreddit lacked “accountability” and “transparency.” After the reddit platform sanctioned the subreddit amid substantial international press coverage, the moderators also invoked the language of governance, making a formal public statement that “the mods directly responsible for this system are no longer a part of the team and the new team is committed to maintaining a transparent style of moderation.” (BBC, 2014; Collier, 2014).

**Internships, Applications, and Elections: Becoming a Moderator on reddit**

The practical work of recruiting and choosing new moderators also requires people to define what it means to be a moderator. Since a subreddit’s current moderators control the reddit software’s process of appointing new moderators, would-be moderators must justify themselves and their ideas of the work to their would-be peers. Likewise, current moderators invest substantial labor into the work of admitting new moderators. At these moments of transition, democratic, oligarchic, and professional notions of moderator work come into tension as subreddits negotiate who should select the leaders and what qualities they should demonstrate.

Among those interviewed, moderators gained their positions through a wide range of means. One was added by a school friend who needed extra help. Others were invited to be moderators after demonstrating substantial participation in the subreddit’s affairs. One was made a moderator in appreciation of their role to exposes the scandal over sexual images of minors. Some were recruited for their expertise at operating the reddit platform software. Yet many subreddits also operate formal structures for adding moderators, systems that draw from the language of the workplace and the public sector.

Many subreddits hold a formal application process for becoming a moderator. In the simplest versions, interested parties fill out an interview form, noting their time zone and availability, describing their moderation experience, listing their skills, and explaining their reasons for applying. One popular subreddit received 600 applications in one recruitment effort, identified a shortlist of 60 applicants to interview, and chose from the shortlist. The process from call to selection can take from weeks to over a month.

While moderator teams sometimes take final responsibility for selecting new moderators—what Shaw and Hill call oligarchy—some subreddits open the final selection to subscribers. The reddit platform doesn’t support ballots, so subreddits have developed their own voting systems. Speaking about elections in a community for people from marginalized groups in the United States, a moderator explained, “I got one ballot, just like every one else.” Yet especially with elections, moderators still felt responsible to filter possible nominees lest the wrong person become elected. The same moderator explained that public opinion wasn’t appropriate for nominating candidates since it risked reinforcing prejudice: “lots of people who can’t be bigots so much anymore [due to social pressure] have found that they can still target [minority group] and nobody seems to mind.”

If voting software supplies infrastructure for democratic notions of moderation, the job board for finding experienced moderators outside of a community offers infrastructure for more oligarchic forms of leadership. This subreddit publishes moderation opportunities alongside “offers to mod.” Postings routinely offer arguments on the nature of moderation work, such as the disinterested approach to moderation offered in one job listing for a community with frequent conflicts:

> I’m looking for an impartial moderator, who doesn’t belong to [organization], and who doesn’t hold a specific view on it. Must have:

- been on reddit for at least 2 years
- moderating experience

The sub is an open platform to discuss [topic], but prejudiced comments aren’t allowed.

Soon after the primary moderator posted this message, community members, who had noticed the listing, added objections: “Seriously? We have posted so many requests for mods to that sub. We have even posted solutions that result in a very balanced 3 party system.” These community members accused the poster of delinquency and argued strongly against the idea of disinterested, objective moderation: “Anyone without knowledge on the subject will be unable to effectively moderate the sub.” After an extended discussion, the moderator accepted their proposal, and the “three party system” was still in place over 1 year later.

Even democratic subreddits emphasize previous experience when selecting moderators, leading many to seek and tout their moderation “résumé.” Since a medium-to-large subreddit is unlikely to accept applicants with limited experience, some subreddits grow their labor pool by offering “internships” and other entry-level moderation opportunities. /r/SubredditOfTheDay, which publishes original content every day, offers a 2-month internship for people seeking moderation opportunities. Interns agree to write six original
posts that feature interviews with the moderation teams of other subreddits. Those who finish the internship period are made full moderators, and they also gain opportunities to moderate other subreddits.

The process of choosing moderators is one of the most powerful ways to define the meaning of moderation and acculturate moderators to that meaning. Even during attempts at democracy or oligarchy, the other stakeholders still shape this acculturation through the platform software, through public pressure, or through the power that moderators have over the process.

**Crises in Legitimacy and the Removal of Moderators**

In technical terms, only two parties can remove a moderator from their position on reddit. Platform employees, known as “admins,” occasionally remove moderators if they are convinced that the moderator was inactive or abusing their power. Moderators with greater seniority also possess the power to remove those within the same community who were appointed more recently.

In an interview, one moderator described a “coup attempt” by moderators who systematically removed others who disagreed with their political views. Someone noticed the attempt in time and reinstated the ejected moderators. In another case, the sibling of someone who moderated a 30,000 subscriber group compromised their reddit account, took charge of the subreddit, and only restored it upon receiving threats of violence. Many moderators, especially those of large or contentious subreddits, pay close attention to their personal information security to protect against such takeovers. Platform employees will also occasionally take action to restore a subreddit’s moderators when asked.

Moderators are more commonly removed for failing to perform their role. In some cases, would-be moderators appeal to the platform, who offer a process for requesting moderation of “inactive” subreddits. In other cases, a moderator loses their legitimacy to govern—as in the case of the technology moderators that were removing all conversations about surveillance. In these cases, community participants sometimes pursue the person they mistrust, incessantly mocking their pronouncements and questioning their decisions. Such cases tend to conclude with a post from the moderator announcing their resignation, or a post from other moderators announcing that the offending moderator has been removed.

**Moderator Compensation and Corruption**

In 2012, a moderator of three of the largest subreddits posted links to an online news outlet after being hired as a social media advisor by the publisher’s marketing firm (Morris, 2012). In response, the reddit platform banned the user and added a rule against third party compensation. Moderators also receive substantial scrutiny and criticism from their communities for alleged “corruption.”

In one case, someone sent messages on the reddit platform to “a few dozen” moderators, offering compensation for help promoting their content. When some moderators reported the offer to reddit, employees investigated the private messages of everyone who received the offer. When the employees noticed that some moderators had responded positively, the company banned their accounts, including moderators of some of the platform’s largest, most popular NSFW subreddits (Martinez, 2013). In 2015, a large gaming company asked moderators to remove links to material that could not legally be published, offering moderators early access to an upcoming Star Wars game in exchange for their help. When one moderator reported the relationship to reddit, the others removed the moderator for a time, until they themselves were banned by reddit for accepting a “bribe.” A reddit representative explained that the gaming company should have used alternative channels to address illegally shared material (Khan, 2015). In another case, a mobile phone manufacturer offered “perks” to moderators of a subreddit that commonly discussed their products. In exchange, the company asked that its employees be made moderators. To protect themselves from community disapproval or platform intervention, moderators reported the request to reddit and posted the offending messages for discussion by their community (Farrell, 2015).

In interviews, moderators were insistent that they did not seek compensation, arguing that news articles that focused on their unpaid status failed to understand the nature of their work. One interviewee brought up the AOL community leader program, arguing that reddit moderators were different because they weren’t managed as closely as the AOL volunteers. This independence was important to many moderators, including one who claimed, “I don’t think I work for reddit. I run communities and reddit is the tool I use to do that.” Yet at the time of the reddit blackout, moderators also felt ignored by the company behind these “tools.” One explained that “it doesn’t help when the site you are on doesn’t appreciate/recognize/care about the cumulative thousands and thousands of hours the mods put in to make their site usable.”

**Starting Subreddits and Governing Moderator Networks**

While some new subreddits are created to support a pre-existing community, many moderators describe “founding” a subreddit and developing a growing community over time. Yet even the work of creating new subreddits requires managing the expectations of platform operators, moderators, and community participants. In interviews, I observed these
negotiations among relationship-themed subreddits and networks of subreddits.

Relationship subreddits offer listings of people who are looking for conversations, penpals, and relationships, sometimes sexual, but often not. When one moderator started a group for users of a mobile messaging system, their goal was to help newcomers on the messaging platform “find more people to chat with,” whatever age. As the subreddit grew, participants continued to post requests for relationships and conversations that could be illegal for minors. These “dirty” relationship requests also put the subreddit at risk of intervention from reddit employees. Rather than designate the subreddit “NSFW,” which would limit minors from accessing the group, the moderator created a parallel subreddit for “dirty” relationship matching. By splitting the conversation, the moderator found a way to meet community expectations while also protecting the primary subreddit from platform intervention. When asked why they moderated a community that wasn’t safe for children, the moderator explained that “I never intended to moderate a NSFW subreddit. It blew me away the community want for it.”

Creators of new subreddits also work to comply with the expectations of other moderators, especially if they seek to join a subreddit “network.” These networks are jointly managed collections of subreddits that share moderators and a common governance structure. Some networks specialize in a particular kind of content. Several offer inspiring general-interest photography; others share celebrity pornography. Some networks adopt a structure akin to city states. To join the network, a moderator must grow their subreddit to a minimum size, institute a set of network-designated policies, and convince a “champion” within the network to advocate for their inclusion. These champions also help new network members comply with the network’s requirements. New subreddits are inducted by vote from the moderators. At the time of writing, the largest two networks included 169 and 117 constituent subreddits, although networks also occur at smaller scales.

One network stopped accepting new subreddits after participants in a newly added subreddit began “doxing” reddit users—a practice of publishing the addresses and phone numbers of people they disliked:

one time we added a sub, vetted them, once we approved them, they started posting information on reddit users, so it looked like [the network] had approved doxing, which was one of the two things that could get us banned [by the company].

Rather than risk reprisals from the platform operator, the network dissociated itself from the offending subreddit and halted all new applications. To address future risks, they required all groups to accept a lead moderator from the network’s central leadership, to keep “everyone pointed in the same direction.”

Acknowledging Moderators’ Position With Platform, Community, and Other Moderators

Two regularly shared comic strips by former moderator Daniel Allen remark directly on the work that moderators must do to manage their relationships with their communities, other moderators, and the reddit platform. The first “life of a mod” comic strip presents moderators as people who carry out a wide range of community care for little appreciation. In the comic, moderators are janitors, referees, police, educators, and artists (Figure 1). The second presents the “Life of a Secret Cabal Mod,” drawing attention to the accusations of oligarchy that moderators receive. The heading of each panel includes a common accusation toward moderators. The illustration beneath each heading offers an alternative explanation for the behavior that attracts accusation. For example, when one moderator helps another learn to remove what they see as hate speech, they could be accused of conspiring to silence dissent. When platform employees share software updates and moderators pass on community complaints to the company, they might also be accused of collusion (Figure 2). By drawing attention to the complicated negotiations that moderators conduct in multiple directions, Allen’s comics themselves make a case for how those parties should see moderators.

Civic Labor in the Reddit Blackout

Scholars of moderation work have rightly identified the stakeholders that moderators face as they negotiate the meaning of the work. This “civic labor” requires moderators to serve three masters with whom they negotiate the idea of moderation: the platform, reddit participants, and other moderators. Moderators differ in the pressure they receive from these parties and the weight they give them. Some face further stakeholders outside the platform. Yet attempts to make sense of moderation by focusing on any one of these relationships can bring the other actors out of focus. These limitations become apparent when attempting to make sense of the reddit blackout, which was not a labor dispute, not always a collective action from communities, and not entirely a coordinated action by a bloc of organized moderators seeking to consolidate power. All three of these interlocutors in the boundary work of moderators are apparent in prior research on the factors that predicted a subreddit’s chance of joining the blackout. Those models show that community-related factors as well as factors in the relations between moderators predicted the likelihood of a subreddit to put pressure on the company (Matias, 2016). Across the population of subreddits, moderators found the decision thrust upon them. Their actions represent the outcomes of unique negotiations with the three parties who together bring their work into being.
Deciding to Join the Blackout

The reddit blackout was precipitated when the company dismissed an employee who had consistently offered direct support to moderators in some of the site’s most popular discussions: live question-answer sessions with notable people, called Ask-Me-Anything threads (Isaac, 2015). Moderators of the /r/IamA subreddit described being caught off guard while in the middle of a live Q&A. When they disabled their subreddit to decide their response (Lynch & Swearingen, 2015), other moderators of large subreddits took note. To these observers, the company’s failure to coordinate the transition with moderators was another sign of its neglect of moderator needs. Moderators had already been attempting to convince the company to improve moderator software and increase its coordination with moderators. In interviews, moderators explained that moderators of the largest groups had previously dismissed the idea of blacking out. But “after she was fired, the idea came up again, [and] no one was really against it.” These moderators described the blackout as a tactic that might give greater leverage to company employees who routinely advocated for moderator interests. When other moderators observed the behavior of these large groups, many joined the blackout, leaving messages on their subreddits expressing “solidarity” for moderators affected by the blackout.

Even as moderators discussed the blackout with each other, they also negotiated pressures from their communities over the decision to join the blackout. In interviews, moderators described receiving large volumes of private messages from participants that urged them toward or against the blackout. In response, many posted discussion threads asking for community opinions or announcing their decisions.
In one post, a moderator apologized for “the inconvenience of going dark” and explained,

I did get messages from people. The more I watched and saw more and more subs going down, I figured it was worth sending a message [to the platform]. We had kind of a mod vote and decided to black out.

Community interests were considered in many moderator decisions. One group of gaming-related subreddits, whose moderators see it as an “island just barely within reddit” concluded that joining the blackout would “punish our users who don’t know or don’t care about reddits politics.” Yet they still faced pressure from many their community to join the blackout: “we eventually released the statement after we received dozens of modmails and posts on both subreddits.”

Some moderators invited their communities to vote on participation in the blackout. In many cases, moderators followed the results of community votes. Yet networks of moderators did not always agree with their communities. In one subreddit in a subreddit network, one moderator held a vote that came out in favor of the blackout. The rest of the network stayed active; moderators more central to the network described the vote as a “rogue faction” and ignored it. Instead, they issued a proclamation that the entire network would stay out of the protest. Elsewhere, one moderator described their community vote as a way to distract those who were clamoring for the blackout, gaining time for moderators to reach a collective decision. Many moderators and participants questioned the legitimacy of the votes that did occur, guessing that the results might be skewed by influxes of reddit users beyond their community who wanted to influence a community’s decision.

Across these situations, moderators faced the same three questions: what would their actions say to the platform, to other moderators, and to their communities? The effect of the blackout on reddit’s civic labor would not be limited to their relationship with the company—it would affect every other relationship in their everyday moderation work.

**Defending Decisions After the Blackout**

Moderators also faced the consequences of their decisions once the blackout concluded. When the platform operators quickly ceded to moderator demands, many declared victory. Community and moderator reactions were more complex. While some subreddits systematically removed any mention of the blackout, it was more common for moderators to post a discussion explaining what had happened. Especially for subreddits that were disabled for the entire weekend, this conversation could be heated. Only a small number of participants might notice a vote called at the moment of decision; many more would feel the effects of a blacked-out community. At these moments, moderators often defended themselves by referring to these votes. “You’re all upset about the blackout decision. Which is silly. If you were upset why didn’t you raise your concerns?” one wrote. In other cases, moderators assigned responsibility to a single moderator acting alone. Sometimes, they offered statements that they removed the person from the moderation team or encouraged them to resign.

In many of these discussions, moderators expressed support for the blackout, explained the reasons one might join the protest, and also apologized to their communities. These statements positioned moderators as supporters of the blackout while also defending themselves from community critiques. One recipe-sharing subreddit moderator took a compromise position by briefly joining the blackout and then re-opening in advance of 4 July US Independence Day parties. They expressed their “full support” for the other moderators, drew attention to an overwhelming community vote to black out, and then wrote an apology: “we are deeply sorry for the outage. Things need to change on reddit, and this was our best way to let them know our demands.”

**Conclusion: Civic Labor Online**

While the details of volunteer moderation are always under negotiation, the negotiations surrounding this civic labor always face platform operators, community participants, and other moderators. Scholarly accounts of moderation are right to draw attention to these different stakeholders, but a clearer account of moderation work should attend to all three at once, just as moderators must always do. All three forces acculturate a moderator to their ever-changing position, from the application process to the moment they step down or are removed.

From the most common dispute over a single comment removal to collective actions that make international news, the meaning of moderation is described in all three ways as people define and redefine the boundaries of moderation. Calling this work civic labor allows us to acknowledge the complex and contingent nature of volunteer moderation throughout the conversations that draw and redraw its meaning together with platforms, the public, and moderators themselves.

These stakeholders are not an exclusive list. For example, during the reddit blackout, two reddit moderators published a *New York Times* opinion article in the attempt to retain their celebrity guests and large public audience (Lynch & Swearingen, 2015). Yet I argue, based on my fieldwork, that negotiations with these three stakeholders are central to any discussion of volunteer governance online.

This civic labor has been a recurring pattern in a 40-year history of volunteers being invited, elected, and chosen into governance positions online. Nor is it unique to for-profit platform; moderators of non-profit platforms such as Wikipedia face a similar set of stakeholders to maintain their roles, as do the journalists involved in fact-checking news on Facebook (Ananny, 2018).
It is possible that civic labor may also be found beyond online platforms: in debates over the unionization of school street-crossing guards, among parents who coach community sports within for-profit leagues, in the elected school boards of publicly funded private schools, or in the everyday governance work of scholarly peer review. In all these cases, volunteers do more than just the work associated with their role: they must negotiate the meaning of their civic role and power with each other and with a wider system that relies on their labor.

Even if civic labor is unique to our digitally mediated social lives, the sense we make of this work will shape our capacity to build meaningful relationships online while protecting public safety, managing our civil liberties, and upholding principles of justice. By recognizing that work more clearly, we can build the understandings we need to address these challenges as a society.

**Acknowledgements**

This work was undertaken while I was a summer intern at Microsoft Research. I owe special thanks to the hundreds of reddit users who participated in this research. I am also deeply grateful to Tarleton Gillespie and Mary Gray for offering mentorship and feedback throughout this research, as well as the Oxford Internet Institute brownbag seminar, who offered generous feedback on an early version of this argument.

**Declaration of Conflicting Interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Funding**

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded as part of an internship at Microsoft Research.

**Notes**

1. [http://web.archive.org/web/20150703012219/http://www.reddit.com/about](http://web.archive.org/web/20150703012219/http://www.reddit.com/about) (accessed 3 July 2015)
2. [http://web.archive.org/web/20150704143845/https://about.twitter.com/company](http://web.archive.org/web/20150704143845/https://about.twitter.com/company) (accessed 4 July 2015)
3. Many accounts have multiple moderator positions, and some use “throwaway accounts” and “alts” on reddit (Leavitt, 2015). While this number is based on an empirical analysis I conducted in June 2015, the number of accounts may be greater than the number of people involved.
4. Quotations from subreddit discussions have been obfuscated to protect participant privacy.
5. [https://www.reddit.com/r/science/comments/43g15s/first_transparency_report_for_rscience](https://www.reddit.com/r/science/comments/43g15s/first_transparency_report_for_rscience)

**References**

Ananny, M. (2018). Checking in with the Facebook fact-checking partnership. *Columbia Journalism Review*. Retrieved from [https://www.cjr.org/tow_center/facebook-fact-checking-partnerships.php](https://www.cjr.org/tow_center/facebook-fact-checking-partnerships.php)

BBC (2014). Reddit downgrades technology community after censorship. *BBC News*. Retrieved from [https://www.bbc.com/news/technology-27100773](https://www.bbc.com/news/technology-27100773)

Bernstein, M. (2003). Collage, composites, construction. In *Proceedings of the fourteenth ACM conference on hypertext and hypermedia* (pp. 122–123). New York, NY: ACM.

boyd, d. (2010). Social network sites as networked publics: Affordances, dynamics, and implications. In *Networked self: Identity, community, and culture on social network sites* (pp. 39–58). London, England: Routledge.

Boyte, H. C., & Kari, N. N. (1996). *Building America: The democratic promise of public work*. Philadelphia, PA: Temple University Press.

Bruckman, A. (1998). Finding ones own in cyberspace. In C. Haynes & J. R. Holmveik (Eds.), *High wired: On the design, use, and theory of educational MOOs* (pp. 15–24). Ann Arbor: University of Michigan Press.

Butler, B., Sproull, L., Kiesler, S., & Kraut, R. (2002). Community effort in online groups: Who does the work and why. In S. P. Weisband (Ed.), *Leadership at a distance: Research in technologically-supported work* (pp. 171–194). Hoboken, NJ: Lawrence Erlbaum Associates.

Centivany, A., & Glushko, B. (2016). “Popcorn tastes good”: Participatory policymaking and Reddit’s. In *Proceedings of the 2016 CHI conference on human factors in computing systems* (CHI’16, pp. 1126–1137). New York, NY: ACM.

Chen, A. (2015). When the Internets moderators are anything but. *The New York Times*. Retrieved from [https://www.nytimes.com/2015/07/26/magazine/when-the-internets-moderators-are-anything-but.html](https://www.nytimes.com/2015/07/26/magazine/when-the-internets-moderators-are-anything-but.html)

Collier, K. (2014) Reddit’s technology has a secret list of about 50 words you can’t use in headlines. *The Daily Dot*. Retrieved from [https://www.dailydot.com/news/reddit-technology-banned-words/](https://www.dailydot.com/news/reddit-technology-banned-words/)

Crawford, K., & Gillespie, T. L. (2014). What is a Flag for? Social media reporting tools and the vocabulary of complaint. *New Media & Society*, 18, 410–428.

Duggan, M. (2014). *Online harassment*. Retrieved from [http://www.pewinternet.org/2014/10/22/online-harassment/](http://www.pewinternet.org/2014/10/22/online-harassment/)

Farrell, N. (2015). *HTC tried to bribe a Reddit moderator and got burned . . . hard*. Retrieved from [https://medium.com/@notarobot/htc-tried-to-bribe-a-reddit-moderator-and-get-burned-hard-b82f68446fae](https://medium.com/@notarobot/htc-tried-to-bribe-a-reddit-moderator-and-get-burned-hard-b82f68446fae)

Geiger, R. S. (2016). Bot-based collective blocklists in Twitter: The counterpublic moderation of harassment in a networked public space. *Information, Communication & Society*, 19, 787–803.

Geiger, R. S., & 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* (pp. 117–126). New York, NY: ACM.

Gieryn, T. F. (1983). Boundary-work and the demarcation of science from non-science: Strains and interests in professional ideologies of scientists. *American Sociological Review*, 48, 781–795.

Gillespie, T. (2010). The politics of platforms. *New Media & Society*, 12, 347–364.

Gillespie, T. (2018). *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*. New Haven, CT: Yale University Press.
