The Civic Labor of Online Moderators

J. Nathan Matias

Abstract

Volunteer moderators of online platforms have done fundamental work to foster social relations for over forty years. Moderators create, support, and control public discourse for millions of people, even as their uncompensated labor upholds platform funding models. In this paper, I examine the "civic labor" of moderators on the social news platform reddit, where a strike by over two thousand subreddit communities in July 2015 forced the company to meet their demands. Scholarship on volunteer moderation has tended to view this work as digital labor, civic participation, or oligarchy. In mixed-methods research with over 52,000 subreddits and over a dozen interviews, I show how the everyday meanings of moderation work are negotiated as moderators face the platform, their communities, and other moderators alike. In disputes over moderator decisions, in the process of choosing moderators, and in the governance of wider networks of many subreddits, moderators must manage their position with all three stakeholders. I also show how the recognition of this civic labor brings clarity to complex moments of collective action like the reddit blackout. Volunteer governance continues to be a common approach to managing social relations, conflict, and civil liberties online. Our ability to recognize the nature of moderation work will shape our capacity to address those challenges as a society.

Introduction

On July 2, 2015, volunteer moderators of over two thousand two hundred "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 one week later, and within two months, the company had hired its first Chief Technical Officer, partly to improve the platform's moderation software (36).

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 a New York Times Magazine article on the word "moderator," Adrian Chen wrote (6):

> 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 (42; 37; 32). Legal theorists and computer scientists describe moderators as civic leaders of online communities who build their own public spheres (22); much of this scholarship outlines general strategies to structure governance work for fair and functional communities at scale (4; 15). A third conversation draws from from the sociology of participation to consider the social structures of those who acquire and exercise moderation power, finding that common tendencies toward oligarchy on platforms like Wikia (40) may be necessary for the survival of online communities (43).

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 offer peer support. In these conversations, especially at moments of tension and transition, moderators negotiate how they are seen by the communities they govern, other moderators, and the operators of the online platform.

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

Corresponding author:

J. Nathan Matias, MIT Media Lab, 75 Amherst St, Cambridge MA, USA, 02139

Email: jnmatias@mit.edu

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.

In their everyday work, moderators must satisfy and explain themselves to all three parties, 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.

This everyday work of defining moderation has implications well beyond community interests or academic debate. 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 (9). Debates over compensation struggle with the challenges of paying this essential labor that many people are willing to do without pay. Since professional services reportedly charged between 4 and 25 cents USD per comment in 2014, active moderation remains unlikely for many large platforms (21). Yet America Online(AOL) community leaders settled a class action lawsuit over unpaid wages for $15 million USD (24). In recent years, many news organizations have disabled public discussions, unable to afford moderation costs (16).

In this paper, I examine the everyday "boundary work" (13) carried out by moderators to negotiate the idea of moderation as they face platforms, communities, and each other. To foreground the ways that moderation work is defined in negotiation with all three parties, I introduce the idea of "civic labor" to describe work that is not solely done for one of these parties, but which is constituted in the act of responding to all three.

Moderation Work

Volunteer moderators have played a fundamental role in social life online for over 40 years, from librarians in 1970s Berkeley looking after local message-boards (3) to today's Facebook group administrators (25), Wikipedia arbitrators (32), and reddit moderators. Although not all work of fostering community is carried out by formal moderators, people in these formal positions are founders, maintainers, content producers, promoters, policymakers, and enforcers of policy across the social internet (4). On many platforms, moderators also manage autonomous and semi-autonomous moderation software that carries out this work (12).

By delegating policy and governance power to moderators, platform operators reduce labor costs and limit their regulatory liability for conduct on their service while positioning themselves as champions of free expression and cultural generativity (14). This governance work invites public scrutiny, which companies often face for their responses to flagged material (8). 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.

The evolution of moderation over the history of reddit 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" one month later (17). Over the next two years, the company started dozens of new subreddits, mostly to separate conversations in different languages. In Jan 2008, after its acquisition by Cond Nast and 10 months after introducing advertising, the company launched "user-controlled subreddits." Before then, 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" (18; 19).

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 one hundred fifty thousand moderator roles[‡] for over fifty-two thousand 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: attempting to theorize the role of their volunteer work in platform business models. Among examples in open source and free culture, this scholarship frequently refers to labor organizing by community leaders (essentially moderators) of AOL chatrooms and other communities in the 1990s. Initially eager to offer this service in exchange for discounts, credit, and other perks, some of the 14,000 "community leads" came to see their work as unpaid labor. Moderators filed a class action lawsuit in 1999, prompting an inconclusive U.S. Department of Labor investigation and eventually winning $15 million dollars from AOL in a 2008 settlement (38; 24).

In her 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]." (42).

In a series of articles on the AOL lawsuit, Postigo explores the nature of the delicate symbiosis between platforms and

---

[*] http://web.archive.org/web/20150703012219/http://www.reddit.com/about Accessed July 3, 2015

[†] http://web.archive.org/web/20150704143845/https://about.twitter.com/company Accessed July 4, 2015

[‡] Many accounts have multiple moderator positions, and some use "throwaway accounts" and "alts" on reddit (26). Consequently, this number over-estimates the number of people involved.

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, 70s, and 80s. 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 re-imagined 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 (37; 38).

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 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 put it in the New York Times. Yet in many ways, the reddit blackout defies explanation by these theories. 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 Centivanny 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 (5).

Moderation as Civic Participation

The work of moderation online is the work of creating, maintaining, and defining "networked publics," imagined collective spaces that "allow people to gather for social, cultural, and civic purposes" (2). 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 his colleagues describe community maintenance as a service to the community itself (4). Consequently, their survey research imagines moderation similarly to any community work. Aside from the unique challenges of tending community software, people 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 during conflicts 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 (11).

Moderation as Oligarchy

Even as their work supports community, the power of moderators is defined and managed by other people who gate-keep the process of becoming a moderator. A third perspective on moderation work examines the interests of moderators when they 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 (39). 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 (40). Where Shaw and Hill see oligarchy, others see experience necessary for online communities to flourish. Also studying Wikia, Zhu and colleagues interpreted similar findings to argue that communities whose leaders also lead other communities are more likely to survive and grow (43). 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.

Standpoint and Methods

I came to this research just after leading a team to study efforts by Women, Action, and the Media (WAM!), an NGO offering support to people experiencing harassment on Twitter (31). 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 work that Twitter should be paying for (33). Others called it a dangerous form of censorship (41). As our team studied the work of reviewing and responding to Twitter harassment, I was deeply moved by the overwhelming amounts of labor and personal risk taken by the harassment reviewers. Volunteers handled cases at all hours and became harassment targets themselves. One volunteer dropped out after experiencing severe post-traumatic stress (31). Furthermore, WAM! also needed to manage their relationship with Twitter to retain the privilege

of supporting harassment receivers and maintain a voice on Twitter's policies.

My fieldwork with reddit moderators began at a time when I was trying understand the many-sided scrunity that WAM!'s harassment reviewers had faced. Volunteer responders might be unpaid, but they were 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 the discursive positioning that 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 four-month period from June through September 2015, with followup data collection through February 2016. I focused on moments of tension and transition, including the process of becoming a moderator, transitions of leadership, conflicts between communities, crises of legitimacy, and collective action during the reddit blackout of July 2015. 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 conducted 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, including those on 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.

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 (29). 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 could either accept their power or 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 themself 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 janitoral 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 towards the reddit company, they disagreed. One described the language of janitor as "a response to complaints about conspiracies, censorship, etc" rather 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, and others publish 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

---

[§]Quotations from subreddit discussions have been obfuscated to protect participant privacy

[¶]https://www.reddit.com/r/science/comments/43g15s/first_transparency_report_for_rscience/

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 (34). In another case, participants accused a large technology subreddit's moderators of censoring political discussions. To support these accusations, one reddit user 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 amidst substantial international press coverage, the moderators made 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." (1; 7)

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

The practical work of recruiting and choosing new moderators 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 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 for their role in exposing 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 timezone 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 the elections in one subreddit for a minority group, a moderator explained, "I got one ballot, just like every one else." Yet especially with elections, moderators still felt responsible to to filter possible nominees lest the wrong person become elected. The same moderator explained why public opinion wasn't appropriate for selecting candidates: "lots of people who can't be bigots so much anymore 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 moderator job board on reddit offers infrastructure for more oligarchic forms. This subreddit publishes moderation opportunities alongside "offers to mod." These postings offer explicit 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 a disinterested moderator from the outside: "Anyone without knowledge on the subject will be unable to effectively moderate the sub." After a substantial discussion, the moderator accepted their proposal, and the "three party system" was still in place over one year later.

Even democratic subreddits emphasize previous experience when selecting moderators, leading many to seek and tout their moderation "résumé." Since a medium or large subreddit is unlikely to accept applicants with limited experience, several subreddits grow their labor pool by offering "internships" and other entry-level moderation opportunities. /r/SubredditOfTheDay, which publishes original content every day, offers a two-month internship for people seeking moderation opportunities. Interns agree to write 6 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.

Among large subreddits that admit inexperienced moderators, newcomers are sometimes admitted in cohorts and offered mentorship that can last several months. As new moderators demonstrate their capabilities, they are given greater moderation powers upon election or appointment. Several large subreddits operate internal promotion structures that formalize responsibilities at each rank and offer documented criteria for career advancement in moderation.

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 who are more junior.

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, 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 (35). 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 (28). In 2015, a large gaming company offered moderators early access to an upcoming Star Wars game if they would remove material that could not legally be published. When one moderator reported the relationship to reddit employees, 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 (23). 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 (10).

In interviews, moderators were insistent that they did not seek compensation, and 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 felt ignored by the company. 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.

> I never intended to moderate a NSFW subreddit.
> It blew me away the community want for it

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.

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 by 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 doxxing, 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 towards moderators. The illustration beneath each heading offers an alternative explanation for the behavior that attracts accusation. 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, they might also be accused of collusion with the company (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 across all their work: the platform, reddit participants, and other moderators. Moderators differ in the pressure they receive from these parties and 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

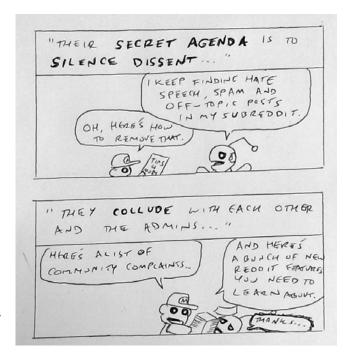**Figure 1.** "Life of a Mod" comic by former moderator Daniel Allen, /u/solidwhetstone

Figure 2. Details from "Life of a Secret Cabal Mod" comic by former moderator Daniel Allen, /u/solidwhetstone

communities, and not entirely a coordinated action by a bloc of organized moderators seeking to increase their power. All three kinds of language are present in debates over the meaning and nature of the blackout. 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.

Civic labor is apparent in the statistical models I have already published on the factors that predicted a subreddit's chance of joining the blackout. Those models show, on average across the roughly fifty-two thousand subreddits active at the time, 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 (30). By examining moderators' blackout decisions and

community reactions after the end of the blackout, I show how moderators managed those pressures.

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 (20). 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 (27), other moderators of large subreddits took note. To these moderators, 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 their 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 or decline participation in 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 the voting process as a way to distract those who were clamoring for the blackout, a way to gain time to discuss a collective decision with other moderators. Many questioned the legitimacy of the votes, guessing that the results might be skewed by influxes of reddit users beyond their community who wanted to influence their decisions.

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 constrained to their relationship with the company—it would affect every other relationship in their everyday moderation work.

Defending Decisions After the Blackout

Moderators 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, 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 July 4th U.S. 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

decision, to collective actions that make international news, the meaning of moderation is defined in all three ways together. Even as Adrian Chen argued in the New York Times Magazine that the blackout was the act of a "moderatocracy," he was acknowledging and engaging in the kinds of civic and platform rhetoric that also shape a moderator's position.

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 arrangements; moderators of non-profit platforms such as Wikipedia face a similar set of stakeholders to maintain their roles. These stakeholders are not an exclusive list. For example, two reddit moderators published a New York Times opinion article during the blackout in the attempt to retain their celebrity guests and large public audience (27). Yet the work of volunteer governance within a larger system is defined in conversation with at least these three.

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 are held accountable for civic power that they also negotiate 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 those challenges as a society.

Acknowledgements

References

[1] Reddit downgrades technology community after censorship. *BBC News*, April 2014.

[2] danah boyd. Social Network Sites as Netowrked Publics: Affordances, Dynamics, and Implications. In *Networked Self: Identity, Community, and Culture on Social Network Sites*, pages 39–58. Routledge, 2010.

[3] Amy Bruckman. Finding ones own in cyberspace. *High Wired: On the Design, Use, and Theory of Educational MOOs. Ed. Cynthia Haynes and Jan Rune Holmevik. Ann Arbor, MI: U of Michigan P*, pages 15–24, 1998.

[4] Brian Butler, Lee Sproull, Sara Kiesler, and Robert Kraut. Community effort in online groups: Who does the work and why. *Leadership at a distance: Research in technologically supported work*, pages 171–194, 2002.

[5] Alissa Centivany and Bobby Glushko. "Popcorn Tastes Good": Participatory Policymaking and Reddit's. In

Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16, pages 1126–1137, New York, NY, USA, 2016. ACM.

[6] Adrian Chen. When the Internets Moderators Are Anything But. *The New York Times Magazine*, July 2015.

[7] Kevin Collier. Reddit's r/technology has a secret list of about 50 words you can't use in headlines. *The Daily Dot*, April 2014.

[8] Kate Crawford and Tarleton L. Gillespie. What is a Flag for? Social Media Reporting Tools and the Vocabulary of Complaint. *New Media & Society*, August 2014.

[9] Maeve Duggan. Online Harassment, October 2014.

[10] Nicholas Farrell. HTC Tried to Bribe a Reddit Moderator and got Burned... Hard, September 2015.

[11] R. Stuart Geiger. Bot-based collective blocklists in Twitter: the counterpublic moderation of harassment in a networked public space. *Information, Communication & Society*, 19(6):787–803, June 2016.

[12] R. Stuart Geiger and David Ribes. 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. ACM, 2010.

[13] Thomas F. Gieryn. Boundary-Work and the Demarcation of Science from Non-Science: Strains and Interests in Professional Ideologies of Scientists. *American Sociological Review*, 48(6):781–795, 1983.

[14] Tarleton Gillespie. The politics of platforms. *New Media & Society*, 12(3):347–364, 2010.

[15] James Grimmelmann. The Virtues of Moderation. SSRN Scholarly Paper ID 2588493, Social Science Research Network, Rochester, NY, April 2015.

[16] Anika Gupta. *Towards A Better Inclusivity: Online Comments and Community at News Organizations*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, April 2016.

[17] Steve Huffman. what's new on reddit: for those of you with a private office..., January 2006.

[18] Steve Huffman. what's new on reddit: brace yourself, ads are coming, March 2007.

[19] Steve Huffman. what's new on reddit: new features, January 2008.

[20] Mike Isaac. Reddit Moderators Shut Down Parts of Site Over Employees Dismissal. *The New York Times*, July 2015.

[21] Justin Isaf. Justin Isaf - How To Reduce Your Moderation Costs, October 2014.

[22] Christopher Kelty. Geeks, Social Imaginaries, and Recursive Publics. *Cultural Anthropology*, 20(2):185–214, May 2005.

[23] Zarmena Khan. EA Reportedly Bribed Star Wars Battlefront Reddit Mods, November 2015.

[24] Lauren Kirchner. AOL Settled with Unpaid "Volunteers" for $15 Million. *Columbia Journalism Review*, February 2011.

[25] Matthew J. Kushin and Kelin Kitchener. Getting political on social network sites: Exploring online political discourse on Facebook. *First Monday*, 14(11), October 2009.

[26] Alex Leavitt. This is a Throwaway Account: Temporary Technical Identities and Perceptions of Anonymity in a Massive Online Community. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 317–327. ACM, 2015.

[27] Brian Lynch and Courtnie Swearingen. Why We Shut Down Reddits Ask Me Anything Forum. *The New York Times*, July 2015.

[28] Fidel Martinez. Top Reddit porn moderators banned for alleged bribes. *The Daily Dot*, January 2013.

[29] Adrienne Massanari. # Gamergate and The Fappening: How Reddits algorithm, governance, and culture support toxic technocultures. *New Media & Society*, page 1461444815608807, 2015.

[30] J. Nathan Matias. 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, USA, 2016. ACM.

[31] J. Nathan Matias, Amy Johnson, Whitney Erin Boesel, Brian Keegan, Jaclyn Friedman, and Charlie DeTar. Reporting, Reviewing, and Responding to Harassment on Twitter. *arXiv:1505.03359 [cs]*, May 2015. arXiv: 1505.03359.

[32] Amanda Menking and Ingrid Erickson. The Heart Work of Wikipedia: Gendered, Emotional Labor in the World's Largest Online Encyclopedia. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 207–210. ACM, 2015.

[33] Robinson Meyer. The Good (and the Bad) of Twitter's New Bid to Stop Harassment. *The Atlantic*, November 2014.

[34] Kevin Morris. Reddit shuts down teen pics section. *The Daily Dot*, October 2011.

[35] Kevin Morris. Reddit moderator banned for selling his influence. *The Daily Dot*, April 2012.

[36] Drew Olanoff. Reddit Names Marty Weiner, Founding Engineer At Pinterest, Its First CTO. *TechCrunch*, August 2015.

[37] Hector Postigo. Emerging sources of labor on the Internet: The case of America Online volunteers. *International review of social History*, 48(S11):205–223, 2003.

[38] Hector Postigo. America Online volunteers. *International Journal of Cultural Studies*, 12(5):451–469, September 2009.

[39] Jennifer Preece and Ben Shneiderman. The reader-to-leader framework: Motivating technology-mediated social participation. *AIS Transactions on Human-Computer Interaction*, 1(1):13–32, 2009.

[40] Aaron Shaw and Benjamin M. Hill. Laboratories of Oligarchy? How the Iron Law Extends to Peer Production. *J Commun*, 64(2):215–238, April 2014.

[41] Andrew Sullivan. The SJWs Now Get To Police Speech On Twitter. *The Dish*, November 2014.

[42] Tiziana Terranova. Free labor: Producing culture for the digital economy. *Social text*, 18(2):33–58, 2000.

[43] Haiyi Zhu, Robert E. Kraut, and Aniket Kittur. The impact of membership overlap on the survival of online communities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 281–290. ACM, 2014.