Fediverse Blocklists: Moderation in Noncapitalist Social Media

Robert W. Gehl

York University, Toronto, Canada, rwg@yorku.ca, www.robertwgehl.org

Abstract: Content moderation is a key form of labour on social media. While much of the scholarly attention has been given to paid or voluntary content moderation on corporate social media, this paper draws attention to content moderation on noncapitalist, alternative social media. Specifically, it focuses on the use of shared instance blocklists on the fediverse, a noncentralised network of community-run social media sites. The paper draws on critical analysis of the act of listing, which finds that listing is an administrative and moral act that introduces three problems: lists don't carry their own selection criteria, they are binary, and they can grow. However, listing also produces knowledge. Drawing on this literature as well as participant observation and interviews, the paper explores how fediverse blocklist developers attempt to mitigate the problems of lists while also generating knowledge about content moderation in noncapitalist social media.

Keywords: content moderation, blocklists, fediverse, alternative social media, prefiguration

1. Introduction

Writing in *The Verge,* Nilay Patel (2022) argues that social media has one main product: content moderation. And of course, every product requires labour to produce. Whether or not content moderation is the *sole* product of corporate social media is debatable. However, it is a key component of the overall social media governance process (Gorwa 2019, 856). The labour of content moderation is a highly exploited element in global corporate social media capitalism. Once hidden "behind the screen" (Roberts 2014), content moderation has drawn the attention of scholars (Gillespie 2018) and journalists (Newton 2019; Pilling and Murgia 2023). As Shahram Azhar writes, big tech corporations "are complicit in the exploitation of a concealed global workforce that experiences... job insecurity, unemployment, and severe mental health issues due to the precariousness of their contracts on the one hand, and the repulsive nature of their work on the other" (2021, 155). When paid content moderators attempt to organise – as contract workers working for Meta in Kenya tried in 2023 – their union efforts are busted (Wangari and Mutandiro 2023). And, given the content these workers are asked to moderate – including scenes of violence and rape, child abuse, and political hate – it is no wonder the workers exhibit symptoms of PTSD (Newton 2019).

Corporate social media do not just rely on paid content moderators. Their work is supplemented by the free labour of volunteers. Anna Gibson has documented the work of Facebook Group volunteer moderators, finding that they see this work as a second job (Gibson 2022; 2023). Similar practices happen on Reddit, where volunteer mods enforce community norms on subreddits (Thach et al. 2024). These contemporary practices reflect what Tiziana Terranova (2000) found in the late 1990s with AOL moderators: the "nearly ubiquitous exploitative structural inequalities of digital platforms" that rely on voluntarily-provided labour to function.

Both paid and unpaid moderators create value for social media corporations. One valuable aspect is that their moderation makes the platforms safe for brands and advertisers, protecting brands from having their advertisements appear alongside lewd or disturbing content (Roberts 2016). Since the collapse of advertising revenue would be disastrous for corporate social media, these firms have developed "Trust and Safety" teams and practices as their "own means to protect themselves—in part, by delineating space, authorizing exchange and fortifying what they accrue in revenue, brand and indistinct 'value'" (Denyer Willis 2023, 1878). This labour provides an infrastructure for what is now called "surveillance capitalism" (Zuboff 2019), a business model where the users create content and social media sells the resulting data – and user attention – to marketers.

Ultimately, however, academic and journalistic critique of corporate social media will do little to change the underlying dynamics of profit-seeking corporations. There has been academic critique of corporate social media for two decades now (e.g., Coté and Pybus 2007; Petersen 2008; Fuchs 2012; Lovink and Rasch 2013). Instead of hoping corporate social media improve, we need solutions. These could include state regulation of corporate social media or publicly-funded social media services. Another approach – the one I will favour here – is to push past capitalist realism (Fisher 2009) to consider existing alternative practices as responses to the criticisms of corporate social media.

We know of the problems of content moderation in capitalist corporate social media. Less is known about content moderation practices on non-capitalist, alternative social media systems. This paper adds to our critical knowledge of non-capitalist, alternative social media content moderation by examining federated social media (or the fediverse). The fediverse is a network of tens of thousands of individual servers or "instances", each with anywhere from one to hundreds of thousands of members. The network is predominantly volunteer-run and not-for-profit (Struett et al. 2024).

As I will show, this network has developed governance practices radically different from corporate social media. One of these will be the focus of his paper: shared instance blocklists. These lists are now a vital governance aspect of the fediverse. The use of shared instance blocklists in non-capitalist, alternative social media shifts "trust and safety" thinking away from the needs of marketers, advertisers, and capitalists towards the needs of communities. This is not only reflected in the sociotechnical elements of shared instance blocklists, but also in the heated debates that the very presence of blocklists inspire.

In what follows, first I will explain my methods. Next, I will draw on theories of prefiguration to examine the fediverse as an actual and potential non-capitalist system with very different governance norms. Then, I will explore content moderation practices on the fediverse in general and the use of shared instance blocklists in particular.

2. Methods

This paper is part of a larger project which began in 2017 and focused on the fediverse and has since resulted in a book (Gehl 2025). The key approach has been participant observation. Much of this involves simply signing up for federated social media accounts and engaging in their affordances: filling out profiles, following other accounts, posting, and liking, boosting, or commenting upon other people's posts. This aspect of participant observation provides a user's perspective on the network.

However, there is also a technical substrate to attend to. Contemporary federated social media relies on a technical protocol, ActivityPub (Lemmer-Webber et al. 2018). ActivityPub is a W3C standard, meaning that it is open to anyone to implement ("W3C Patent Policy" 2020). Likewise, many of the ActivityPub implementations (e.g., Mastodon, PeerTube, GoToSocial) are free and open source software (FOSS) projects. Thus, unlike corporate social media, federated social media allows for greater inspection and modification of the underlying technical infrastructure, from how data are stored to how they are moved around the network. My participant observation therefore includes running federated social media software, both in experimental settings and in support of online communities.

In addition to the technical aspects, running federated social media as an administrator has exposed me to social practices, such as content moderation, guiding new users, writing policies, and how the server I administer relates to other servers. Thus, my participant observation combines the technical and social practices of users, administrators, and content moderators.

I have supplemented participation observation with interviews, both for the larger project and for this paper specifically. Because consent is highly valued on the fediverse (Pincus 2024; Roscam Abbing and Gehl 2024), my approach to interviews draws on online research ethics guides and seeks informed consent from interviewees (frantze et al. 2019). For the larger book project, I have interviewed over 40 fediverse admins, moderators, organisation representatives, and users. For this particular paper, I draw on several of those interviews as well as specific interviews with shared blocklist developers.

3. The Fediverse as Prefigurative, Non-capitalist Space

My larger fediverse project has been an attempt to heed Gibson-Graham's call for scholars to "perform new economic worlds, starting with an ontology of economic difference – 'diverse economies'" (Gibson-Graham 2008, 614). In my work, I have been pushing past what we might call "surveillance capitalist realism", instead exploring non-capitalist, community-run alternatives to corporate social media (Gehl 2015; 2017). Other scholars have examined alternative social media. Many focus on the fediverse (Rozenshtein 2024; Frost-Arnold 2024; Gow 2022; Mansoux and Roscam Abbing 2020; Theophilos 2024). Others have examined systems such as Secure Scuttlebutt (Mannell and Smith 2022). This line of scholarship links alternative social media to larger histories of alternative media, a form of media that should "not only be understood as alternative media practices, but also as critical media that question dominative society" (Fuchs 2010, 174). Specifically, a major question posed by alternative social media centres on sociotechnical systems of governance, including content moderation. Alternative social media scholarship draws on critiques of corporate social media practices – including the critiques aimed at corporate governance and content moderation – to explore and contribute to alternatives.

Thus, the field of alternative social media scholarship has the potential to take part in media activism which attempts to build a media environment outside of surveillance capitalist realism. "Prefiguration" has been a useful – if contested – conceptual label for this kind of practice (Yates 2021; Maeckelbergh 2011; Yates 2015). "Prefiguration is a different kind of theory, a 'direct theory' ... that theorises through action, through doing" (Maeckelbergh 2011, 3). Prefiguration is "a way in which protest is carried out and as a movement's building of alternatives" (Habersang

2022, 3). It is a smaller-scale implementation of the world activists are striving to build on a global scale.

As an alternative social media system, the fediverse's prefiguration includes a range of sociotechnical elements. A key element is the licensing of the code, typically as FOSS. This enables anyone with the technical ability to inspect, modify, and install fediverse software code on an instance, which in turn runs on an open source protocol to communicate with other instances. This in turn cedes some degree of control to end users.

However, FOSS is not enough to make for a non-capitalist network. FOSS licensing, for example, stems from less labour-conscious cultures (Ross 2006, 747).[1] And there is no guarantee that a given fediverse instance won't be run in the same manner as Instagram or X, with the data generated by users being sold "to advertising clients, who in return for paying money get targeted access to users' profiles that become advert spaces" (Fuchs 2024, 134). Fediverse instances also use governance practices that differentiate them from surveillance capitalism, such as the use of codes of conduct instead of exploitative terms of service agreements (Gehl and Zulli 2022). Codes of conduct function as both internal moderation documents, but also as a means to decide how to federate with other instances. A common prohibition among fediverse instances is against marketing, advertising, and branding practices – indeed, if a fediverse server is installed that does intend to sell user data, that server would very likely be blocked by the bulk of the network. Thus instead of serving the needs of advertisers, fediverse instances are more likely to serve the needs of their members.

Alongside the use of FOSS and codes of conduct, many fediverse instances are run as nonprofits, either informally or formally. Informal nonprofits include instances where the admin donates the service to the community, sometimes asking for donations. Formal nonprofits are organised according to the regulations of the state in which the admin lives (Kissane and Kazemi 2024). Server costs (and, in some cases, payments to moderators) are underwritten by donations in a form more akin to mutual aid (Spade 2020) than commodity exchange, and many of these nonprofit instances are also run with informal or formal democratic processes, such as co-ops.

In addition, fediverse admins and communities often take a degrowth-oriented stance on network development (Kwet 2022), purposely building communities rather than simply trying to grow big user counts. This is also tied to governance: as my interviewees often noted, some instances close registrations to new users in order to keep their communities management for content moderation (Zulli, Liu, and Gehl 2020, 1196). While there are certainly fediverse advocates who want a bigger network – including those who welcome Meta's Threads, an ActivityPub-enabled corporate social media site – a common ethos of the fediverse is to keep instances small so as to maintain local governance norms as well as avoid having nodes in the network become too powerful.

As such, the fediverse reflects the three strategic components of prefiguration Yates (2021) identifies: reproduction, mobilisation and coordination. Reproduction involves creating spaces of care where resources required for action are maintained – e.g., a daycare centre for activists with childrearing responsibilities. In the case of

---

[1] However, some fediverse projects use non-FLOSS, ethical licenses that do explicitly push back against capitalism. BookWyrm, for example, uses an anti-capitalist license ("BookWyrm" [2020] 2023).

the fediverse – as will be particularly clear in the following sections – admins and moderators work to maintain spaces where members can socialise without having to deal with surveillance capitalism or misogynist, racist, or transphobic harassers. Mobilisation involves "processes of deploying social movement resources" (Yates 2021, 1046). It is the exercise of movement power. This is the most common connotation of "prefiguration": mobilisation is the act of creating the alternative space in the first place. The activists who create fediverse software often do so precisely for this reason – to create a non-capitalist, community-run alternative to corporate social media. Coordination, finally, is "imagining, marshalling, planning and guiding of forces in a particular direction or towards a particular end" (Yates 2021, 1046). On the fediverse, this is reflected in the quip "the favorite topic of discussion on the fediverse is the fediverse." That is to say, fediverse members – including academics – critique the fediverse, debating and contributing to its development with a larger goal of subverting the surveillance capitalist system (if not capitalism itself) (Mansoux and Roscam Abbing 2020, 137).

While critiques of prefigurative projects, such as alternative forms of media, note they "tend to idealise small-scale production and tend to neglect orientation towards the political public" (Fuchs 2010, 174), the fediverse functions as a global network of many small communities, helping push prefiguration to the more effective approach of "oscillating between, on the one hand, experimentation and the building of 'alternative' ways of living and relating, with attempts to consolidate and proliferate their outcomes on the other" (Yates 2015, 2). While small fediverse instances may experiment locally with their codes of conduct and governance practices, they also tend to expand outwork through the ActivityPub protocol, connecting to other like-minded instances around the world in a global political formation I have elsewhere called the "covenantal fediverse" (Gehl 2025).

As I will show in the next section, content moderation practices are vehicles for such local-to-global oscillations, placing the instance-as-community at the centre of the social imaginary of the fediverse. Shared instance blocklists are a particularly complex part of fediverse content moderation – this will be my focus throughout the remainder of the article.

4. Governance and Content Moderation on the Fediverse

As should be clear, fediverse instances are often governed in a radically different manner than corporate social media. Simply put, corporate social media are centralised, with power accruing to the owners. The fediverse, by contrast, is comprised of tens of thousands of self-governing communities. But, much as in the case of corporate social media, a key product these communities produce is content moderation, or processes through which good communication is distinguished from bad.

Content moderation on the fediverse involves multiple, overlapping practices. Individuals are often given many tools for personal content moderation. On Mastodon, for example, individuals can mute or block other individuals or entire instances, and they can create lists of terms that will be prevented from appearing on their timelines. Another content moderation practice is found at the instance level. Each instance has an admin who handles the technical infrastructure. In some instances, the admin doubles as a content moderator, although instance members can be promoted to that role. Whether done by admins or moderators, content moderation often involves writing a code of conduct and thus setting the rules of the instance and enforcing those rules if members of the instance violate them.

Both the individual and instance-level moderation practices can face inward toward the instance itself: individuals can block other individuals who have accounts on the same instance, for example, or moderators can silence individuals on the instance who violate the community's code of conduct. However, according to my participant observation and interviews with fediverse members, such internal-facing moderation practices are rare. This is likely because members who join an instance self-select, agreeing to abide by that instance's local code of conduct.

Indeed, much of the stressful and involved content moderation occurs during instance-to-instance relations, what Kissane and Kazemi (2024) call "federated diplomacy". While admins and moderators have a great deal of control over their own instances, by design they have no control over the membership of other instances – that is how non-centralised federation works. This leads to a problem: what can be done when someone on another instance harasses people on my local instance? As one interviewee told me, the ease with which a fediverse server can be spun up means that small groups of internet trolls can set up their own servers and harass people relatively easily: "You buy a cheap, cheap domain name, you spin up a cheap, cheap server, put Pleroma [software for federated microblogging] on it, boom, 20 minutes later, you're calling people the n-word". Or, an individual who gets blocked for harassment can sign up for a new account on another server and begin harassing people again. Moderating these events is an outward-facing, instance-to-instance practice.

Fediverse members and admins have created two tools to combat these issues. One involves shared hashtags. #Fediblock, for example, is a hashtag created by two feminists on the fediverse, Marcia X and Ginger, to publicly share knowledge about harassing accounts (X 2024). People would use #fediblock to warn others about trolls who migrated from one instance to another. Similar hashtags have also been used to coordinate blocking of alt-right platforms, such as Gab (Caelin 2022).

Inspired by #fediblock and similar hashtags, several people have developed publicly shared instance blocklists. "Shared bocklists in use today usually take the form of lists of servers, or sometimes lists of users, that meet some threshold for being bad actors in the Fediverse as defined by the people who maintain the blocklist" (Kissane and Kazemi 2024, 64). This can combat the problem of "cheap cheap" services coming online: rather than asking individual instance admins to block new, problematic servers one-by-one, admins can now import blocklists created by other admins, blocking problematic instances en masse either through importing a file or subscribing to an automated service.

5. The Problems and Promises of Blocklists

However, as Kissane and Kazemi found, blocklists are seen as "a necessary but flawed tool" (Kissane and Kazemi 2024, 64). In their analysis of fediverse governance, which drew on interviews with fediverse admins, they report that "no admins we spoke to were unreservedly positive about shared blocklists" (Kissane and Kazemi 2024, 59). This comports with my years of participant observation on the fediverse, where I have observed the ambivalence many admins, moderators, and fediverse members have towards shared instance blocklists. Blocklists are seen as necessary because they solve major problems in instance-to-instance relations – namely, how to quickly and effectively sever connections between instances that do not share values. They are flawed for many reasons. Here, I will explore the

problems of fedivese blocklists, drawing on scholarship on blocklists specifically as well as the practice of listing in general.

Recent scholarship on social media blocklists echo the current fediverse admins' mixed feelings about instance blocklists. In their study of Twitter blocklists, particularly those associated with #GamerGate, Jhaver et al. (2018) identify problems such as false positives and blocklist proliferation (where blocklists are copied and used without any auditing). Scholars have also identified problems with blocklists in digital media, particularly email spam and cybersecurity blocklists: "despite their wide adoption, many open-source blocklist providers lack clear documentation about their structure, curation process, contents, dynamics, and interrelationships with other providers" (Feal et al. 2021, 1334). These problems lead to the same ones identified by Jhaver et al: false positives and proliferation.

But the problems of lists go deeper than the digital age. Fundamentally, the problem stems from the fact that listing is an act of power – those who list claim some authority to decide what counts. When we add in the fact that blocklists list groups of people, we are talking about fundamental questions of governance, whether on the fediverse or in general (Bowker and Star 1999, 137). Scholars have raised concerns about any form of moral listing, such the Black Lists of the Reformation of Manners campaigns in 18th century England (Hurl-Eamon 2004) to the Hollywood blacklist of the mid-20th century US red scare (Humphries 2010) to social credit in China (Trauth-Goik and Liu 2023). Fediverse blocklists, along with spam lists or Twitter account lists, arguably fall into the same broad category of these forms of moral lists – they are "designed to identify untrustworthy actors, or those otherwise deemed unacceptable to the makers of the list, for penalty or restriction" (Trauth-Goik and Liu 2023, 1021).

Complicating matters is the fact that all lists are both reductive and expansive. Lists are a very rudimentary form of writing, with no grammar (Delbourgo and Müller-Wille 2012, 711; Helton 2019). They do no more than include or exclude (Tankard 2006) – either something is on the list or it is not (Goody 1977, 106). They are metonymical at best, abstracting from social context (Garson 1996). Crucially, this means that lists *do not include their own selection criteria* (Jensen 2011; Feal et al. 2021). Thus, although it is a power-knowledge instrument of control across time and space, the list itself is incapable of demonstrating why something is listed – that is the domain of some third party. This raises a major problem for moral lists: how did something get listed? The list itself will not reveal the answer.

Lists are also expansive. While lists are bounded (Goody 1977, 81), they can be added to, seemingly infinitely (Delbourgo and Müller-Wille 2012). And while we may think the digital age is the age of copying and proliferation, scholars who examined the ancient lists note that those lists were copied again and again. This raises major questions about listing, including moral listing: when does it end?

However, listing is more than a reductive marker of moral disapprobation: lists are also administrative tools that help produce knowledge (Helton 2019) and coordinate activity across time and space (Bowker and Star 1999, 138; Geiger 2016). As anthropologist Jack Goody (1977) documents, lists mark the transition from oral to

written cultures, particularly in ancient Sumeria and Assyria. These lists were deeply tied to the administrative state (Delbourgo and Müller-Wille 2012, 711). Goody argues that the lists of this period included three broad types. There were records of events, roles, situations, persons – "a kind of inventory of persons, objects, or events." There were something analogous to shopping lists, or guides "for future action, a plan." And another is the lexical list, *Listenwissenschaft* – lists of concepts, a proto-dictionary (Goody 1977, 80).

Lists, including moral lists, can thus be productive and generative. For example, Laura Helton's research celebrates Black American listmakers, the "bibliographers, collectors, and library workers who managed Black archives in the early twentieth century" (Helton 2019, 83). Such lists helped produce a field of Black literature instead of burying it under the racist analyses produced by white authors. More recently, blocklisting has been shown to produce valuable knowledge. In his study of Twitter blocklists, R. Stuart Geiger (2016) argued that the construction, governance, and application of blocklists contributed to "collective sensemaking" among the Twitter users who relied upon them, including knowledge about what entails harassment and what to do about it.

To gloss the problems of lists, then, there are three. First, lists do not include their own selection criteria. Second, they reduce things to a binary choice – things are either on-list or off-list. Third, lists can expand and proliferate. However, these problems are all deeply wrapped up in a benefit of lists: they contribute to knowledge among communities of people, including moral knowledge. This is why lists are dangerous practices of administrative control. For whom is the knowledge produced? What are the effects of this knowledge? Who benefits from this control? This brings us to the heart of the ambivalence about shared instance blocklists on the fediverse: the lists themselves require governance.

In the next two sections, I will draw on documents, participant observation, and interviews with shared instance blocklist administrators to show how these actors are navigating the governance problems of shared instance blocklists. While this includes potential pitfalls of listing, it also includes the production of shared community knowledge that makes non-centralised, instance-focused fediverse governance distinct from that of corporate social media.

6. Mitigating the Problems of Shared Blocklists on the Fediverse

Overall, instance blocklist creators and administrators are deeply aware of the problems of lists. They have also developed practices to mitigate those problems.

6.1. Lists do not Contain Criteria

Lists themselves do not contain their selection criteria. When it comes to shared instance blocklists, for those who disagree with their use – and for those who find their instances listed – the blocklist itself provides no indication of why this instance was listed while that instance was not. Often, then, criticism of the instance blocklist falls on whomever developed the list. My interviewees faced harsh criticism for developing their lists. Ro, who developed The Bad Space blocklist, told me, "A lot of the contention comes from, 'Oh, well, who are you who gets to decide who goes on

that list?'". Oliphant, who created one of the earliest instance blocklists, said that when he started publishing his list, "what I found out, right away, is that this is extremely threatening" to "bigots". And Jaz-Michael King, the founder of a nonprofit called IFTAS, which provides a list called CARIAD, noted that blocklisting is "such a charged topic".

To shift criticism away from the personality of whomever built the list, these blocklist operators write lengthy documents explaining the processes by which instances are added. IFTAS, for example, produces an extensive library of documents, including the CARIAD Policy, with inclusion criteria and a discussion of possible biases of the source materials ("CARIAD Policy" 2024). IFTAS also publishes audit logs. The Bad Space also posts selection criteria ("About" 2024). As for Oliphant, those lists are published to open source code repositories.

However, while publishing criteria provides para-documentation to the lists, the existence of selection criteria documents do not themselves mitigate the criticism that these lists are the products of individual people. Recognising this, all of the lists are generated by collectives, not individuals. As Ro said, "a lot of people have brought up governance, and governance is an important piece in terms of deciding who gets to decide who is going to be on that list". Collective governance of the blocklists can also mitigate another problem of lists: they are binary, on/off documents.

6.2. Lists are Binary

List operators also acknowledge the fundamental binary aspect of listing. As powerful as listing is, listing is also seen as crude and reductive. Either something is listed or it is not. In the case of blocklists, those instances that are listed are blocked by whomever imports the list, and this "disconnects everything", Ro told me. "In terms of the connections that people have built, the history there, and things like that. There is a concern with that". This severity and reductiveness is perhaps the most common criticism of blocklists – even blocklist creators echo the criticism. As Jaz-Michael King of IFTAS declared, "blocklists are a *terrible* tool. It is the hammer in the toolbox. We just have very few tools currently".

However, as noted in the previous section, all three blocklists (the Bad Space, Oliphant's Lists, and the IFTAS's CARIAD) are not the products of individuals. As Ro said, "The Bad Space is a huge collaboration behind the scenes. I keep saying this: people keep trying to make it about me and my proclivities as far as who gets blocked. I'm like, no, man! This stuff is coming from an amalgamation of sources that have been doing this work for a very long time". The "amalgamation of sources" come from other admins (10 as of this writing). The Bad Space compiles their specific instance blocklists into "heat ratings" – the greater percentage of sources blocking a specific instance, the higher the rating. For those who want to use a list from The Bad Space, they can set their "heat" tolerance level and import a .CSV file reflecting the heat rating they choose. Oliphant's list and IFTAS's CARIAD also work the same way. None of these lists are the products of individuals, but are the products of consensus across many fediverse admins.

The result of this are not simple, binary lists, but multiple lists with spectra produced through varying degrees of collective consensus. The Bad Space, Oliphant's List, or IFTAS's CARIAD are each multiple lists made by many people. The most stringent list, Oliphant's Tier 0, requires only two admins to have blocked an instance for that instance to appear. However, a would-be blocklist importer may not want that level of strictness – they may opt instead for a tier with a higher tolerance. The Bad Space's Heat Rating system is similar, effectively producing

multiple lists. Thus, rather than being singular, binary, on-or-off lists, the collectively produced blocklists are effectively collections of lists, each with different criteria. The goal is to offer a would-be importer options ranging from blocking major portions of the fediverse to only blocking a few. (And of course, not using a list is also an option.)

6.3. Lists can Expand and Proliferate

The third major problem of lists is that they can expand and proliferate. While being finite, any list can be added to. Lists can easily be copied – even the ancient stone tablets anthropologists and historians have studied show evidence of copying and proliferation (Goody 1977). In the case of moral lists, such as blocklists, the fear is that anyone can be added over time and that the lists will propagate across the network.

To mitigate the expansion problem, all of the lists under study offer appeals procedures. These range from informal to formal and even automated. In terms of informal appeals, Oliphant noted that the process is

> You make amends. You go to the one or two or three or however many servers blocked you and say, "Hey, is there anything I can do?" But you have to make real amends…. If you can make the case to a few servers – like for example, 8 out of 10 did block a server, you just go to one server, get them to agree to remove the block. Now you're at 7 of 10. You immediately get off Tier 0, move to Tier 1. So you'd still be on a blocklist, but you could move yourself down a severity ranking just by making amends to one server.

However, Oliphant argues it's rare for people to do this work. Instead, he notes, they often double-down on whatever behaviour got them listed. "You can always tell who a person is when they get blocked", he told me.

The Bad Space has a more formal procedure. It publishes a form and describes a process where the collective will vote on an appeal after interviewing the admin who requested it ("Appeals" 2024). IFTAS also has a formal appeals procedure for its blocklist ("CARIAD Policy" 2024). IFTAS is also developing tools to automatically remove blocks. As Jaz-Michael King told me,

> We want to be able to retract a block. Errors will happen, right? That's a given. And if we are willing to take on the mantle of saying we bear responsibility for advising you on severing human connection, we feel we have the duty of care to advise you to reverse things when we see a change.

Such a system requires a would-be blocklist importer to subscribe to IFTAS's CARIAD for continual updates – including removal of listed instances – rather than downloading a .CSV file and importing it. The Bad Space is also exploring such automation.

As for proliferation, a fear among blocklist critics is that the existence of easy-to-import blocklists means that these lists will be replicated across the network. In cases of mistaken or unjust blocks, those instances that are blocked would have little chance to connect to the rest of the network. The mitigation the blocklist organisations suggest is human intervention. Oliphant's introduction to his blocklist page notes, "it's assumed you are a human being with agency, and take responsibility for the blocks on your server, no matter how they get there, either via

blocklist import or being entered manually" (Oliphant 2022). Ro notes The Bad Space is opt-in. As Jaz-Michael King told me,

> I'm always very cautious to remind folks: if you do not like what we're doing, do not use it. I do not feel that IFTAS wields anywhere near the influence people fear it may such that we would be the new, centralized source of something. We work strongly to preserve polycentric approaches. We support and admire all the other work in this space.

To mitigate proliferation, the authors of the blocklists themselves do not advocate for the uncritical use of lists. They echo the findings of Kissane and Kazemi: lists are a flawed tool, to be used with caution (Kissane and Kazemi 2024, 64). Moreover, the software interfaces of the fediverse reinforce this caution: in my experience importing blocklists for the instances I administer, the system warns me about severing connections and provides data on what specific connections will be cut. I have to affirmatively opt-in.

7. The Product of Shared Blocklists

In spite of the challenges, list developers and administrators put in the labour in order to produce a safer fediverse. Since lists are knowledge-making tools, what sort of fediverse is produced through shared instance blocklists?

This study of fediverse blocklists echoes what Geiger found in his study of blocklists on Twitter:

> In studying the historical development of several different blockbots over time, I have seen how such systems are continually developed and redeveloped as people come to better understand what it even means to ask and answer questions like 'Who is and is not a harasser?' and 'What ought to be done about harassment?' ... These systems are ongoing accomplishments in sensemaking, just as Bowker and Star note in their studies of classification systems about race, health, and labor (Geiger 2016, 797).

Ro echoed this in our interview:

> We don't know what instances are out there that have bad intentions. It's kind of like a grinding process. When someone says something to you, you have to go research and look at that space, and that exposes you to all kinds of nonsense. So then, you have to talk about it, and then you have to determine what constitutes inappropriate behavior.

In both cases, then, mid-2010s Twitter blocklists or contemporary fediverse instance blocklists require groups of people to determine what harassment is and do so outside of the dictates of corporate media. The Bad Space and IFTAS in particular rely on collectives making sense of harassment and making decisions under their own governance.

However, the Twitter blocklists Geiger studied differ from fediverse instance blocklists in several key ways that illustrate how fediverse governance differs from that of corporate social media. Twitter blocklists were developed by third parties using the Twitter API. They were not central to Twitter, but more ad hoc, and changes at Twitter's centre could – and did – undermine them. In 2023, Block Party stopped supporting Twitter/X due to increased costs of using the API (Weatherbed

2023). Later, Twitter/X changed its blocking implementation, eroding the effectiveness of blocking (CNET Staff 2024). This shifted the power of content moderation back to the centre – Twitter/X itself makes decisions about who should be removed from the service, and these decisions centre the needs of brands and advertisers over the needs of Twitter/X users.

Moreover, Twitter/X blocklists were individual-centric. They focused on blocking individual accounts and were used by individuals. Just as in corporate social media more generally, the relationship remains that of an individual using a corporate service, rather than communities setting their own moderation standards.

Even when Twitter blocklists were viable, the volunteers who created them were not remunerated by Twitter. Thus, Twitter blocklists were based on the hyperexploited labour of volunteers who face long hours and trauma. "Social curation is not a trivial activity. It requires many volunteer users contributing several hours each week and coordinating with one another to moderate sexist, racist, and homophobic content. Reviewing rape and death threats, violent images, and aggressive threats over a long period can be psychologically damaging" (Jhaver et al. 2018, 22). The benefits of this freely-given blocklist administration accrued more to Twitter itself than to blocklist developers.

In contrast, fediverse instance blocklists shift the focus of blocking on communities with the benefits accruing to communities over a corporation. Fediverse blocklists are not individual-centric: they focus on instances, which are communities of practice. The instances that are blocked face this sanction because, as a whole, they do not moderate against hate speech and fascism or surveillance capitalist practices. Unlike post-Musk Twitter, there are no central mechanisms to override these blocks in favour of corporate needs.

Conversely, the instances that avoid being listed tend to do so because they have explicitly developed anti-fascist and non-capitalist codes of conduct that seek to protect their members from threats of violence or corporate surveillance. These communities of practice develop their moderation standards and build their connections to other like-minded communities of practice. As Theophilos finds, networks of instances echo older practices of federation, such as those used by the Zapatistas. This model is "the 'Fedifam' model—a structure that would enable a better network of trust among Fediverse instances. The Fedifam structure, also known as a 'caracole,' is based on the approaches of the Zapatista Army of National Liberation (EZLN)" (Theophilos 2024, 8). This is what a colleague and I have called a "covenantal federalist" approach, where small communities band together through shared ethical values (Gehl and Zulli 2022). The production of instance blocklists is part of this process – communities of practice know their values in part by declaring what values they don't hold. As Ro told me, shared instance blocklists can become "an idea to galvanize around".

Although shared instance blocklists accrue their benefits to instances-as-communities, there is one inescapable problem. Developing blocklists can be traumatising. As the quote from Ro above indicates, researching blocklists requires moderators to be exposed to "all kinds of nonsense" – child abuse, transphobia, and racism, to name a few things. While corporate social media will exploit poorly paid content moderators and offer little in the way of support to them (even firing such workers when they seek to unionise), fediverse admins band together in mutual support. IFTAS provides forums and support services for moderators, and both The Bad Space and Oliphant's Lists grew out of ad-hoc associations of like-minded content moderators and administrators who support one another. As Donatella Della

Ratta argues, such practices are a form of collective ethics of care. "Ethics of care are characterised by an understanding of the self that is relational rather than individualistic, placing human and intimate relations at the centre of life, in stark opposition to 'legalistic contractual thinking, so favored in traditional analyses' which might in fact 'alienate persons, rather than draw them together'" (Della Ratta 2020, 111). While the trauma of moderation is likely inescapable, its effects do not need to be borne individually – they can be borne collectively.

8. Conclusion

Through their labour, instance blocklist administrators help contribute to knowledge production on the fediverse – knowledge that is necessary for governance and content moderation. This includes, of course, knowledge of how harassment operates on the network and where it is located, but also knowledge on how small instances-as-communities can develop mutual values without resorting to a centralised, surveillance capitalist social media system. Shared instance blocklists are thus part of the larger prefigurative politics of federated, non-capitalist, non-centralised alternative social media. While lists are inherently problematic tools, they are nonetheless powerful ways to develop knowledge and coordinate across networks.

References

"About". 2024. The Bad Space. 2024. https://tweaking.thebad.space/about.

"Appeals". 2024. The Bad Space. 2024. https://tweaking.thebad.space/appeals.

Azhar, Shahram. 2021. The Conditions of the Global Digital Working Class: The Continuing Relevance of Friedrich Engels to Theorising Platform Labour. *tripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society* 19 (1): 154-170. https://doi.org/10.31269/triplec.v19i1.1217.

"BookWyrm". (2020) 2023. Python. BookWyrm. https://github.com/bookwyrm-social/bookwyrm/blob/e9f26b7d50fda5c1cd28d8a41bac5dafd01aeecc/LICENSE.md.

Bowker, Geoffrey C., and Susan Leigh Star. 1999. *Sorting Things Out: Classification and Its Consequences*. Inside Technology. Cambridge, Mass.: MIT Press.

Caelin, Derek. 2022. Decentralized Networks vs The Trolls. In *Fundamental Challenges to Global Peace and Security: The Future of Humanity*, edited by Hoda Mahmoudi, Michael H. Allen, and Kate Seaman, 143-168. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-79072-1_8.

"CARIAD Policy". 2024. IFTAS Connect. 30 July 2024. https://connect.iftas.org/library/iftas-documentation/cariad-policy/.

CNET Staff. 2024. Blocking on X/Twitter Doesn't Work Anymore. Here's How to Lock Down Your Posts. CNET. 6 November 2024. https://www.cnet.com/tech/blocking-on-xtwitter-doesnt-work-anymore-heres-how-to-lock-down-your-posts/.

Coté, Mark, and Jennifer Pybus. 2007. Learning to Immaterial Labour 2.0: MySpace and Social Networks. *Ephemera* 7 (1): 88-106.

Delbourgo, James, and Staffan Müller-Wille. 2012. Introduction. *Isis* 103 (4): 710-715. https://doi.org/10.1086/669045.

Della Ratta, Donatella. 2020. Digital Socialism Beyond the Digital Social: Confronting Communicative Capitalism with Ethics of Care. *tripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society* 18 (1): 101-115. https://doi.org/10.31269/triplec.v18i1.1145.

Denyer Willis, Graham. 2023. 'Trust and Safety': Exchange, Protection and the Digital Market–Fortress in Platform Capitalism. *Socio-Economic Review* 21 (4): 1877-1895. https://doi.org/10.1093/ser/mwad003.

Feal, Álvaro, Pelayo Vallina, Julien Gamba, Sergio Pastrana, Antonio Nappa, Oliver Hohlfeld, Narseo Vallina-Rodriguez, and Juan Tapiador. 2021. Blocklist Babel: On the Transparency and Dynamics of Open Source Blocklisting. *IEEE Transactions on Network and Service Management* 18 (2): 1334-1349. https://doi.org/10.1109/TNSM.2021.3075552.

Fisher, Mark. 2009. *Capitalist Realism: Is There No Alternative?* Winchester, UK: Zero Books.

frantze, aline shakti, Anja Bechmann, Michael Zimmer, and Charles Ess. 2019. Internet Research: Ethical Guidelines 3.0. Association of Internet Researchers. https://aoir.org/reports/ethics3.pdf.

Frost-Arnold, Karen. 2024. Beyond Corporate Social Media Platforms: The Epistemic Promises and Perils of Alternative Social Media. *Topoi* 43: 1557–1568. https://doi.org/10.1007/s11245-024-10102-2.

Fuchs, Christian. 2010. Alternative Media as Critical Media. *European Journal of Social Theory* 13 (2): 173-192. https://doi.org/10.1177/1368431010362294

Fuchs, Christian. 2012. The Political Economy of Privacy on Facebook. *Television & New Media* 13 (2): 139-159. https://doi.org/10.1177/1527476411415699.

Fuchs, Christian. 2024. Critical Theory Foundations of Digital Capitalism: A Critical Political Economy Perspective. *tripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society* 22 (1): 148-196. https://doi.org/10.31269/triplec.v22i1.1454.

Garson, Marjorie. 1996. I Would Try to Make Lists: The Catalogue in Lives of Girls and Women. *Canadian Literature*, no. 150, 45-63.

Gehl, Robert W. 2015. The Case for Alternative Social Media. *Social Media + Society* 1 (2). https://doi.org/10.1177/2056305115604338.

Gehl, Robert W. 2017. Alternative Social Media: From Critique to Code. In *The SAGE Handbook of Social Media*, edited by Jean Burgess, Alice E Marwick, and Thomas Poell, 330-352.

Gehl, Robert W. 2025. *Move Slowly and Build Bridges: Mastodon, the Fediverse, and the Struggle for Democratic Social Media*. Oxford: Oxford University Press.

Gehl, Robert W., and Diana Zulli. 2022. The Digital Covenant: Non-Centralized Platform Governance on the Mastodon Social Network. *Information, Communication & Society* 0 (0): 1-17. https://doi.org/10.1080/1369118X.2022.2147400.

Geiger, R. Stuart. 2016. Bot-Based Collective Blocklists in Twitter: The Counterpublic Moderation of Harassment in a Networked Public Space. *Information, Communication & Society* 19 (6): 787-803. https://doi.org/10.1080/1369118X.2016.1153700.

Gibson, Anna D. 2022. 'My Other Job': Volunteer Content Moderation as Platform Labor. Dissertation, Stanford, CA: Stanford University. https://www.proquest.com/docview/2786883042.

Gibson, Anna D. 2023. What Teams Do: Exploring Volunteer Content Moderation Team Labor on Facebook. *Social Media + Society* 9 (3). https://doi.org/10.1177/20563051231186109.

Gibson-Graham, J.K. 2008. Economies: Performative Practices for 'Other Worlds'. *Progress in Human Geography* 32 (5): 613-632. https://doi.org/10.1177/0309132508090821.

Gillespie, Tarleton. 2018. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media.* Yale University Press.

Goody, Jack. 1977. *The Domestication of the Savage Mind*. Cambridge, UK: Cambridge University Press.

Gorwa, Robert. 2019. What Is Platform Governance? *Information, Communication & Society* 22 (6): 854-871. https://doi.org/10.1080/1369118X.2019.1573914.

Gow, Gordon A. 2022. Turning to Alternative Social Media. In *The SAGE Handbook of Social Media Research Methods*, edited by Anabel Quan-Haase and Luke Sloan. London: SAGE. https://doi.org/10.4135/9781529782943.

Habersang, Anja. 2022. Utopia, Future Imaginations and Prefigurative Politics in the Indigenous Women's Movement in Argentina. *Social Movement Studies* 0 (0): 1-16. https://doi.org/10.1080/14742837.2022.2047639.

Helton, Laura. 2019. Making Lists, Keeping Time: Infrastructures of Black Inquiry, 1900-1950. *Against a Sharp White Background: Infrastructures of African American Print*, 82-108.

Humphries, Reynold. 2010. *Hollywood's Blacklists: A Political and Cultural History*. Edinburgh University Press. https://www.jstor.org/stable/10.3366/j.ctt1r2bh5.

Hurl-Eamon, Jennine. 2004. Policing Male Heterosexuality: The Reformation of Manners Societies' Campaign Against the Brothels in Westminster, 1690-1720. *Journal of Social History* 37 (4): 1017-1035.

Jensen, Casper Bruun. 2011. Making Lists, Enlisting Scientists: The Bibliometric Indicator, Uncertainty and Emergent Agency. *Science & Technology Studies* 24 (2): 64-84. https://doi.org/10.23987/sts.55264.

Jhaver, Shagun, Sucheta Ghoshal, Amy Bruckman, and Eric Gilbert. 2018. Harassment and Content Moderation: The Case of Blocklists. *ACM Trans. Comput.-Hum. Interact.* 25 (2): 12:1-12:33. https://doi.org/10.1145/3185593.

Kissane, Erin, and Darius Kazemi. 2024. Findings Report: Governance on Fediverse Microblogging Servers. *Manubot*. Manubot. https://fediverse-governance.github.io/.

Kwet, Michael. 2022. The Digital Tech Deal: A Socialist Framework for the Twenty-First Century. *Race & Class* 63 (3): 63-84. https://doi.org/10.1177/03063968211064478.

Lemmer-Webber, Christine, Jessica Tallon, Erin Shephard, Amy Guy, and Evan Prodromou. 2018. ActivityPub. 23 January 2018. https://www.w3.org/TR/activitypub/.

Lovink, Geert, and Miriam Rasch, eds. 2013. *Unlike Us Reader: Social Media Monopolies and Their Alternatives*. Amsterdam: Institute of Network Cultures.

Maeckelbergh, Marianne. 2011. Doing Is Believing: Prefiguration as Strategic Practice in the Alterglobalization Movement. *Social Movement Studies* 10 (1): 1-20. https://doi.org/10.1080/14742837.2011.545223.

Mannell, Kate, and Eden T. Smith. 2022. Alternative Social Media and the Complexities of a More Participatory Culture: A View From Scuttlebutt. *Social Media + Society* 8 (3). https://doi.org/10.1177/20563051221122448.

Mansoux, Aymeric, and Roel Roscam Abbing. 2020. Theses on the Fediverse and the Becoming of FLOSS. In *The Eternal Network*, edited by Kristoffer Gansing and Inga Luchs, 124-40. Amsterdam: Institute of Network Cultures.

Newton, Casey. 2019. The Secret Lives of Facebook Moderators in America. The Verge. 25 February 2019. https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona.

Oliphant, The. 2022. Oliphant. Social Mastodon Blocklists. The Oliphant. 15 November 2022. https://writer.oliphant.social/oliphant/the-oliphant-social-blocklist.

Patel, Nilay. 2022. Welcome to Hell, Elon. *The Verge*. 28 October 2022. https://www.theverge.com/2022/10/28/23428132/elon-musk-twitter-acquisition-problems-speech-moderation.

Petersen, Soren Mork. 2008. Loser Generated Content: From Participation to Exploitation. *First Monday* 13 (3). https://firstmonday.org/ojs/index.php/fm/article/view/2141/1948.

Pilling, David, and Madhumita Murgia. 2023. 'You Can't Unsee It': The Content Moderators Taking on Facebook. *Financial Times*, 18 May 2023, sec. The Big Read. https://www.ft.com/content/afeb56f2-9ba5-4103-890d-91291aea4caa.

Pincus, Jon. 2024. Eight Tips About Consent for Fediverse Developers. The Nexus Of Privacy. 15 April 2024. https://privacy.thenexus.today/consent-for-fediverse-developers/.

Roberts, Sarah T. 2014. *Behind the Screen: The Hidden Digital Labor of Commercial Content Moderation*. University of Illinois at Urbana-Champaign.

Roberts, Sarah T. 2016. Commercial Content Moderation: Digital Laborers' Dirty Work. In *Commercial Content Moderation: Digital Laborers' Dirty Work,* edited by Safiya Umoja Noble and Brendesha M. Tynes, 147-160. Bristol: Peter Lang.

Roscam Abbing, Roel and Robert W. Gehl. 2024. Shifting Your Research from X to Mastodon? Here's What You Need to Know. *Patterns* 5 (1). https://doi.org/10.1016/j.patter.2023.100914.

Ross, Andrew. 2006. Technology and Below-the-Line Labor in the Copyfight Over Intellectual Property. *American Quarterly* 58 (3): 743-766.

Rozenshtein, Alan Z. 2024. Moderating the Fediverse: Content Moderation on Distributed Social Media. In *Media and Society After Technological Disruption*, edited by Justin (Gus) Hurwitz and Kyle Langvardt, 177-192. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781009174411.019.

Spade, Dean. 2020. Solidarity Not Charity: Mutual Aid for Mobilization and Survival. *Social Text* 38 (1 (142)): 131-151. https://doi.org/10.1215/01642472-7971139.

Struett, Thomas, Aram Sinnreich, Patricia Aufderheide, and Robert W. Gehl. 2024. Can This Platform Survive? Governance Challenges for the Fediverse. *International Journal of Communication* 18 (0): 22.

Tankard, Paul. 2006. Reading Lists *Prose Studies* 28 (3): 337–360. https://doi.org/10.1080/01440350600975531.

Terranova, Tiziana. 2000. Free Labor: Producing Culture for the Digital Economy. *Social Text* 18 (2): 33–58.

Thach, Hibby, Samuel Mayworm, Daniel Delmonaco, and Oliver Haimson. 2024. (In)Visible Moderation: A Digital Ethnography of Marginalized Users and Content Moderation on Twitch and Reddit. *New Media & Society* 26 (7): 4034–4055. https://doi.org/10.1177/14614448221109804.

Theophilos, Jamie A. 2024. Closing the Door to Remain Open: The Politics of Openness and the Practices of Strategic Closure in the Fediverse. *Social Media + Society* 10 (4). https://doi.org/10.1177/20563051241308323.

Trauth-Goik, Alexander, and Chuncheng Liu. 2023. Black or Fifty Shades of Grey? The Power and Limits of the Social Credit Blacklist System in China. *Journal of Contemporary China* 32 (144): 1017-1033. https://doi.org/10.1080/10670564.2022.2128638.

"W3C Patent Policy." 2020. W3C. 15 September 2020. https://www.w3.org/policies/patent-policy/#sec-Requirements.

Wangari, Stephanie, and Kimberly Mutandiro. 2023. The Man Leading Kenyan Content Moderators' Battle against Meta. Rest of World. 12 December 2023. https://restofworld.org/2023/kenya-content-moderators-battle-meta/.

Weatherbed, Jess. 2023. Block Party's Anti-Harassment Service Is Leaving Twitter. *The Verge*. 31 May 2023. https://www.theverge.com/2023/5/31/23743538/block-party-hiatus-twitter-app-anti-harassment-service-api.

X, Marcia. 2024. #Fediblock, a Tiny History. Artist Marcia X. 5 November 2024. https://subscriptions.boricua.style/fediblock-a-tiny-history-2/.

Yates, Luke. 2015. Rethinking Prefiguration: Alternatives, Micropolitics and Goals in Social Movements. *Social Movement Studies* 14 (1): 1-21. https://doi.org/10.1080/14742837.2013.870883.

Yates, Luke. 2021. Prefigurative Politics and Social Movement Strategy: The Roles of Prefiguration in the Reproduction, Mobilisation and Coordination of Movements. *Political Studies* 69 (4): 1033-1052. https://doi.org/10.1177/0032321720936046.

Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. First edition. New York: PublicAffairs.

Zulli, Diana, Miao Liu, and Robert Gehl. 2020. Rethinking the 'Social' in 'Social Media': Insights into Topology, Abstraction, and Scale on the Mastodon Social Network. *New Media & Society*, 22 (7): 1188-1205. https://doi.org/10.1177/1461444820912533.

About the Author

Robert W. Gehl
Robert W. Gehl is a Fulbright scholar and award-winning author whose research focuses on contemporary communication technologies. He received his PhD in Cultural Studies from George Mason University in 2010. Before joining York University as an Ontario Research Chair of Digital Governance for Social Justice, he previously held an endowed research chair at Louisiana Tech. He has published over two dozen articles in journals such as *New Media & Society*, *Communication Theory*, *Social Media + Society*, and *Media, Culture and Society.* His books include *Reverse Engineering Social Media*, which won the Nancy Baym Book Award from the Association of Internet Researchers, *Weaving the Dark Web*, and *Social Engineering*, published in 2022 by MIT Press. He also has published a co-edited collection of essays, *Socialbots and Their Friends*.