

Boundary Control as Gatekeeping in Facebook Groups

Malinen, Sanna

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Empfohlene Zitierung / Suggested Citation:

Malinen, S. (2021). Boundary Control as Gatekeeping in Facebook Groups. *Media and Communication*, 9(4), 73-81.
<https://doi.org/10.17645/mac.v9i4.4238>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:

<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more information see:

<https://creativecommons.org/licenses/by/4.0>

Article

Boundary Control as Gatekeeping in Facebook Groups

Sanna Malinen

Department of Social Research, University of Turku, Finland; E-Mail: sanna.malinen@utu.fi

Submitted: 28 February 2021 | Accepted: 21 May 2021 | Published: 21 October 2021

Abstract

Facebook groups host user-created communities on Facebook’s global platform, and their administrative structure consists of members, volunteer moderators, and governance mechanisms developed by the platform itself. This study presents the viewpoints of volunteers who moderate groups on Facebook that are dedicated to political discussion. It sheds light on how they enact their day-to-day moderation work, from platform administration to group membership, while acknowledging the demands that come from both these tasks. As volunteer moderators make key decisions about content, their work significantly shapes public discussion in their groups. Using data obtained from 15 face-to-face interviews, this qualitative study sheds light on volunteer moderation as a means of media control in complex digital networks. The findings show that moderation concerns not just the removal of content or contacts but, most importantly, it is about protecting group norms by controlling who has the access to the group. Facebook’s volunteer moderators have power not only to guide discussion but, above all, to decide who can participate in it, which makes them important gatekeepers of the digital public sphere.

Keywords

Facebook; Facebook groups; gatekeeping; moderation; platforms; political discussions; social networks

Issue

This article is part of the issue “Media Control Revisited: Challenges, Bottom-Up Resistance, and Agency in the Digital Age” edited by Olga Dovbysh (University of Helsinki, Finland) and Esther Somfalvy (Research Centre for East European Studies at the University of Bremen, Germany).

© 2021 by the author; licensee Cogitatio (Lisbon, Portugal). This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

1. Introduction

Social media platforms provide users with vibrant spaces for public discussions across a wide range of topics. By giving individuals the power to network independent of institutions, social media increases collective action and accountability in society (Bennett & Segerberg, 2012; Dutton, 2009; Gustafsson, 2012; Kushin & Kitchener, 2009). In particular, Facebook groups have been identified as a significant arena for citizen engagement as they allow discussion of common interests and goals (Park et al., 2009), and group identity and self-efficacy to be built in relation to participation (Gustafsson, 2012). Currently, Facebook hosts a large number of politically motivated user groups created by political actors and civil activists that reach wide audiences (Gustafsson, 2012; Park et al., 2009; Warren et al., 2014). Previous studies have pointed out that groups on Facebook can promote societal change and provide users with a

channel for expressing counter-discourses to the dominant public voice (Gachau, 2016; Pruchniewska, 2019; Sormanen & Dutton, 2015). However, there is a darker side to social media, and research has pointed out the harmful effects of such platforms on political, economic, and social life, particularly due to the widespread dissemination of misinformation and hate speech through them (Bessi et al., 2016; Del Vicario et al., 2016; Gagliardone et al., 2015).

Social media platforms have the power to promote, delete, and hide content produced by users, making them an important means to shape public discussion (Gillespie, 2018; Gorwa, 2019; Myers West, 2017). Even though they were originally created for facilitating social activity between people and increasing the circulation of user-generated content, they need to be moderated to keep discussions civil and law-abiding. Previous studies have suggested that users are unaware of the moderation policies of platforms and the underlying logic

of these, and these policies are intentionally being kept guarded in order to maintain a sense of openness and freedom that any form of moderation and content control would typically be strongly against (Gillespie, 2018; Roberts, 2016). As Roberts (2016) argued, social media companies want to give their users the impression that content appears on the site simply “in some kind of natural, organic way” (p. 9), and therefore they intentionally obscure human decision-making processes behind moderation. In other words, commercial content moderation is successful when it is invisible as it is not intended to leave any traces (Gillespie, 2018; Roberts, 2016).

Platforms use various moderation strategies for promoting and discarding content. Recently, algorithmic moderation and so-called “filter bubbles” (Pariser, 2011) have received plenty of scholarly attention; but still, less is known about how user-driven modes of content control are organized. Kalsnes and Ihlebæk (2021) argued that user-driven moderation should be viewed as political because volunteer moderators choose to give visibility to some views while hiding others. Particularly when moderation decisions lack transparency, they have serious consequences for participation in public debate (Kalsnes & Ihlebæk, 2021). So far, the obscure nature of social media platforms has made it difficult to study their moderation systems thoroughly (Jhaver et al., 2019; Langlois et al., 2009).

Facebook groups are local, user-created groups hosted by a global platform, which makes their moderation structure complex. Groups’ founders can create and enact their own governance policies, and therefore the rules and moderation practices of individual groups vary greatly throughout the platform. This study focuses on one—and perhaps the most visible—aspect of how these groups are moderated: volunteer moderators who make decisions about acceptable content on a daily basis. It investigates how these moderators create and enact moderating philosophies as intermediaries between group members and the platform. Since not much is known about the work of volunteer moderators, it is important to shed more light on how they shape the visibility of political views in networked social spaces. This study relies on network gatekeeping theory introduced by Barzilai-Nahon (2008) and looks into social dynamics between the stakeholders in the political communities on Facebook groups.

2. Theoretical Background

2.1. Content Moderation

Ever since the emergence of online communities, the way these are moderated has mostly been the responsibility of their founders and key members (Kalsnes & Ihlebæk, 2021). Over the years, scholars have debated whether volunteer moderation is emotionally demanding and labor-intensive unpaid work conducted for the benefit of the companies that run the platforms (e.g.,

Terranova, 2000) or an organic part of community management and development (Seering et al., 2019). Moderators have a key role in determining which content is published and what is removed, and these decisions shape our public discourse (Gillespie, 2018; Jhaver et al., 2019). According to Kalsnes and Ihlebæk (2021), the role of moderators has recently grown to become even more important with the growing prevalence of uncivil online behavior, such as harassment and hate speech that poses a threat to democracy.

Major social media companies have developed moderation strategies for monitoring user-generated content on platforms. However, volunteer users are still the most effective moderators because they understand group norms, are strongly committed to their communities, and derive personal meaning from their moderation work (Gillespie, 2018; Seering et al., 2019). Prior research has shown that volunteer moderators tend to engage personally in the moderation process and view it as a means of growth for both themselves and their communities (Seering et al., 2019). When moderation decisions are left to algorithms or paid moderation teams, communities and their human moderators miss opportunities for guiding discussion and reflecting the values behind it (Ruckenstein & Turunen, 2020; Seering et al., 2019).

User communities hosted by social media platforms differ from traditional, self-governing online communities in terms of their structure. On platforms such as Facebook or Reddit, users can create their own sub-groups and develop specific local policies alongside the platforms’ site-wide rules and terms of use. This complex approach to policy is well exemplified in Facebook groups, whereby users navigate between Facebook’s own community norms, multiple individually tailored, community-devised rules, and implicit cultural codes of conduct. Operating in a multi-layer system with rules derived from a range of sources can be confusing for users (Fiesler et al., 2015). The rules of local groups not only vary across groups, but they can also be rather vague. According to Dovbysh (2021), rules of moderation are individually constructed by group owners who imitate journalistic practices, although they lack professional norms, as was found in the study on Russian Vkontakte groups.

Facebook groups combine both commercial governing mechanisms developed by the platform and self-governance by group members. These two modes of moderation are clearly different in terms of their impact on a group’s social dynamics. Volunteer moderators work from the bottom up, whereas commercial content moderation is directed from the top down, obeying policy and norms set by company. A study by Seering et al. (2019) showed that while company-driven moderation strategies view anti-normative behavior as something that should be removed or banned, volunteer moderators tend to personally engage with community members and view such interaction as an opportunity for growth for the whole community. Some scholars have emphasized this continuous interaction as a

centerpiece for community development and conceptualized volunteer moderation as an ongoing negotiation in which the meaning of moderation is continuously defined and explained amongst stakeholders such as the platform, community, and fellow moderators (Gillespie, 2018; Matias, 2019; Seering et al., 2019). This implies that community guidelines are not fixed and can evolve over time as a result of a company's self-perception and the demands of users; in other words, they are consequent to the negotiation process (Myers West, 2017).

Prior studies have identified fairness as a key element of successful moderation as users' reaction to moderation is likely to depend on whether they feel it is done fairly (Jhaver et al., 2019; Myers West, 2017). If there is confusion about the reasons for moderation or feelings of being treated unfairly, users who have experienced moderation can become frustrated. One way for them to deal with this frustration and confusion is by developing their own theories for content takedown (Jhaver et al., 2019; Myers West, 2017). In particular, hidden commercial content moderation creates tensions between users and the platform: Frustrated users may turn against platforms through collective protests with the aim of raising the visibility of content that the platform has hidden from them (Gillespie, 2018; Myers West, 2017). In the commercial moderation system, users remain absent, and they are only given the role of laborers who can report content they deem objectionable (Myers West, 2017).

2.2. Gatekeeping in Social Networks

For decades, scholars of media and communication have applied a theory of gatekeeping to describe content selection in the media environment and ascribed the term "gatekeeper" to persons who have a role in carrying out this selection. Barzilai-Nahon (2008) has addressed the need for updating traditional gatekeeping theories to fit better in the context of digital networks. According to her, traditional theories view gatekeeping as a selection process based on a gatekeeper's individual characteristics and position of power, while dynamics and relationships between stakeholders are left unconsidered. This reduces gatekeeping to simply a one-way direction and top-down process, which is an inadequate way to describe it in the context of information networks with multiple gates and channels for spreading information (Barzilai-Nahon, 2008). In the context of this study of groups on Facebook, this complexity of information flow is seen in volunteer moderators' ability to control only information within their own communities. Network gatekeeping theory presents three main goals: "locking-in" of gated users inside the gatekeeper's network; protecting established communities from unwanted entry from outside; and maintaining ongoing activities within network boundaries without disturbances (Barzilai-Nahon, 2008). All three goals point to outsiders being the main threat for communities.

Contrary to the traditional literature, which has conceived the gatekeeper as having complete power in information production and dissemination, Barzilai-Nahon (2008) saw a dynamic relationship between the gatekeeper and the gated that forms through frequent, enduring, and direct exchange. The gated are not viewed as passive nor powerless in this process; they too can have power and exercise it. Contrary to traditional media settings, non-elite members can become prominent in gatekeeping in the networked context as well, influencing what is being discussed and how it is done. This is particularly evident in cases of mass movements and uprisings, when ordinary users play a significant role in raising topics to prominence and elevating others to higher status through active gatekeeping (Meraz & Papacharissi, 2013).

Collaborative and networked modes of action were expected to lead to flatter and less hierarchical organizational forms (Bennett & Segerberg, 2012). However, there is evidence that power structures of user-driven communities can be rather oligarchic, so that some individuals gain a more privileged position and exert their authority onto others (Keegan & Gergle, 2010; Shaw & Hill, 2014). As shown by a study of Wikipedia, elite users are in a position to select and remove content, but they also have to accept contributions from non-elite users in order to keep content flowing (Keegan & Gergle, 2010). Similarly, there are elite users with privileges to restrict others in software wiki communities, although they can hinder community development when they use this authority to promote their own agendas over the interests of the community as a whole (Shaw & Hill, 2014). Scholars have presented a range of views on participatory structures of online communities and how these structures are associated with moderation (Keegan & Gergle, 2010; Matias, 2019; Seering et al., 2019; Shaw & Hill, 2014). The key questions here are how should privileged members exert their power over ordinary members, and should they restrict some users to maintain the harmony of the community? In order to survive, online communities need to self-regulate, members must conform to norms by monitoring their own behavior, and those who violate these norms should be punished (Honeycutt, 2005).

The network gatekeeping theory recognizes that the stakeholders involved in gatekeeping are not equally powerful, and some attributes, such as political power, information production ability, or relationship with the gatekeeper, can lead to greater salience in the network (Barzilai-Nahon, 2008). As Myers West (2017) argued, visibility is the most effective way to gain political power on social media in a networked environment, and users without the power to influence platforms' moderation policy can fight unfair moderating decisions by giving prominence to what platform has hidden.

The starting point of this study is that social media users are not just passive receivers of information; instead, they can actively construct their political

environment on social media by building networks and tailoring their information flows. Relying on network gatekeeping theory and its two main components, network gatekeeping identification and network gatekeeping salience, this study aims to investigate gatekeeping practices and goals in political Facebook groups (RQ1) and analyze power dynamics between the gatekeepers and the gated (RQ2). The term “salience” refers to the degree to which gatekeepers give priority to the gated. Therefore, this study examines if there are any differences in members’ positions of power (RQ3), so that some group members are more influential and therefore gain more visibility for their views than others manage to do.

3. Method and Data

This qualitative study uses data obtained from 15 semi-structured interviews with Facebook group moderators as research data. The face-to-face interviews were conducted between December 2019 and February 2020 in Finland. The informants were selected first by searching for active Finnish Facebook discussion groups labeled as political or societal. Then persons named as moderators or administrators of these groups were identified through each group’s public page and contacted personally via Messenger. Initially, interview requests were sent to 20 individuals, of whom five either declined or did not see the invitation. The interviews were recorded and transcribed, and the duration of the voice files varied from 58 to 170 minutes.

This study uses semi-structured interviews because of the flexibility of this format. In addition to predetermined research themes, it allows other relevant themes to develop throughout the interviews (Choak, 2012). Therefore, it can bring out new and unexpected results and allow the study to take new directions. Data analysis followed thematic analysis, which is a process of identifying patterns and themes within the data (Creswell, 2013). The analysis first focused on gaining a detailed understanding of moderation practices and the interviewees’ experiences of their roles. Following the procedure described by Creswell (2013), the text was first classified into codes and then into broader themes. Each of the themes were interpreted in terms of their meaning in respect to the research questions.

This study focuses on groups dedicated to political discussion for three reasons. First, previous research showed that tensions between users of social media networks tend to arise particularly when discussion is connected to politics (Zhu et al., 2017). Second, political beliefs and attitudes are found to drive selectivity in subsequent information processing (Taber & Lodge, 2006). Third, in a political context, information control reflects the state of power relations between stakeholders who aim to achieve their political goals (Barzilai-Nahon, 2008). Prior work has thus suggested that content selection and moderation are very likely to occur in politically

motivated social media discussions in comparison with other topics.

4. Findings

4.1. Moderation as Boundary Control

“You don’t want someone stupid at a good party” (Interviewee 6). As shown in the quote, the interviewees indicated that screening applications for group membership is an essential aspect of moderation. Many moderators reported having developed specific checklists to evaluate who would become a suitable member and contributor and who is applying to the group just to troll. The moderators put a lot of effort into keeping their groups closed from potential troublemakers who might disrupt discussion and prevent other members from participating. In many groups, member lists were curated so that moderators could judge each applicant before giving approval. Sometimes, they would discuss the merits of acceptance with their fellow moderators. Judgement of suitability for the group was passed by inspecting information on an applicant’s profile page and, in particular, their liked pages and other group memberships. Membership of some strictly moderated Facebook groups were perceived as a recommendation and proof of an applicant’s good behavior. As the following quote shows, some moderators were adept at detecting potential trolls and troublemakers by looking for certain signs:

They can be very discreet. Once there was someone who had created numerous troll accounts and each account had the same background picture. But you don’t see that until you put the profiles side by side to compare them. It’s a dog whistle for the like-minded; an invitation to troll. When they see those certain signs in the profile, they will join in the trolling. (Interviewee 10)

The groups being studied were at different stages in their life cycle, which in turn affected their member approval policy. Some groups were rather new and at a growth stage with plenty of applications for membership coming in, and so moderators would accept new members daily. In these groups, the moderators did not scrutinize applications as carefully because they wanted new members to join. Moreover, some groups were at a stage of saturation, and the moderators were satisfied with current membership levels and user activity. In these groups, they were a little reluctant to accept new members and stated that they did not want the group to grow any bigger because new members would bring increased workload and the potential of trouble.

For mature and well-functioning groups, newcomers pose a greater risk as they might challenge existing norms and express their disagreement. In this way, they need extra moderation and guidance. They may be perceived as a threat to the power and authority

of established members, particularly in situations when many newcomers are accepted at the same time (Honeycutt, 2005). However, for online communities to sustain themselves and grow, new members must still be occasionally accepted. One of the groups in this study was at a terminal stage with diminishing user activity and rare new applicants. In this group, moderation policy was very strict and only a few newcomers were accepted. Eventually, this led to the group diminishing.

Similar findings about boundary control have been made in prior research of online communities. In particular, elite members control access to a community and monitor who is allowed to participate in conversations (Honeycutt, 2005; Weber, 2011). If a newcomer fails to conform to group norms, the elite refuse to accept that member into the group unless they admit their ignorance of the norms (Weber, 2011). However, disruption caused by newcomers can be useful for a community as it helps moderators and members to identify and define rules and boundaries. Moderators screen members because they want to minimize damage and avoid additional work. As a moderator from a well-functioning and stable group said: "I admit that when I judge someone's suitability as a member of the group, I think about the potential workload. If their profile information gives the impression that this is a quarrelsome person, I might not approve them" (Interviewee 3).

When members are carefully screened and their views are seen to be similar to the group consensus, moderators are more likely to apply softer moderation strategies. One way of conceptualizing moderation strategies is to divide these into soft and hard, based on how much moderation restricts users' activities in the group. Personal discussions, in which a moderator contacts the member privately and notifies them about questionable behavior, were mentioned as the softest form of moderation, whereas excluding a member from the group either temporarily or permanently was generally considered to be the hardest form of moderation. The strength of moderation can also be defined based on its visibility to users. In this sense, screening members in advance is a soft form of moderation: When someone is not accepted, this does not leave any trace for group members to see because outsiders are not allowed to post in the group. However, declining someone's right to participate can be considered the strongest limitation that a moderator can apply to users. Warning someone discretely in person and hiding someone from discussions without their knowledge are invisible forms of moderation, whereas public interventions in a discussion, bans, and removals are usually visible to the whole group, and therefore might harm one's reputation within the group. Private discussion between moderator and member was considered a discrete form of moderation because it is invisible to other members and allows the person in question to save their face in the group. Private discussion was used particularly in situations when a troublemaking user was well known in the group, or when moderators

suspected that a user might regret their behavior afterwards, for example due to drunkenness. As one moderator said: "In the moderation business, I often feel that we need to protect people from themselves, like preventing them for causing harm to themselves" (Interviewee 6).

Many moderators reported using hard forms of moderation, namely bans and removals, actively and without previous negotiation. However, doing so is likely to cause unwanted reactions in the group as removing someone can create tension and criticism among group members. Because hard moderation is visible to other members, banning or kicking someone out of the group publicly is likely to give even more visibility to their opinions. Removing members tends to invoke critical discussion about censorship and sympathy for the removed person. Moderators admitted that trolls sometimes use this approach to cause extra fuss and reaction from others, and therefore they intentionally provoke moderators with the aim of being punished publicly. For this reason, moderators need to be careful in how they deal with provocative content:

Some people just want to get to say that "the moderation sucks in this group, I am leaving now." And then others begin to wonder if there is something wrong with this group. We have to remove the "I'm leaving" notes because they are used with the intention of harming the group and the good spirit between people. (Interviewee 10)

The findings point out that in successful groups, moderators have a strong sense of community and belonging to their group, which is an important factor behind them actively volunteering for moderation work. Moderators with strong feelings of belonging to their group feel ownership and are committed to taking care of and nurturing the community by continually monitoring content and membership. If founders and moderators do not feel any ownership or obligation to look after their group, it is more likely to become filled with arguments and misinformation. Some moderators mentioned cautionary examples of abandoned, non-moderated groups that attracted political actors who used them to spread their own political agendas. Eventually, the group would drift away from its original purpose.

4.2. Power Dynamics Between Moderators and Members

Another main aim of this study is to understand the power dynamic between moderators and group members, and to find out whether some members' views are given more priority than others. In the interviews, moderators were asked if they perceive comments from all group members as being equal in terms of their value and contribution to the group. They uniformly stated that some social media users are better in having their opinions heard and accepted by the group, while others'

opinions remain ignored. They also described the key characteristics of an influential group member, saying that such a person's individual skills are the most important reason for salience. Asked what most important qualities are for being taken seriously by others, the moderators emphasized good writing and argumentation skills and sound knowledge and expertise of matters under discussion. They also stressed that merely being vocal and active in the group does not make someone influential. If a participant is not good at expressing their opinions in written form or lacks grammatical skills, it is harder for them to be perceived as credible in an online discussion. In addition to skills, being famous through offline activity and having a strong reputation based on previous history as a member also contribute to a member's salience. It seems that a member's personal friendship with the moderator is not perceived as a factor that leads to greater visibility and salience; but instead, these active key members tend to develop closer relationships with moderators and gain more influence over time and activity in the group. These key members have an important role in directing discussion as their opinions are more valued and trusted than those of less-known members. Some moderators admitted that it is difficult to moderate these salient members, and as a result, they are given more freedom to express their views.

Becoming a prominent member who is valued in the group leads to a virtuous circle. Those who are active and comment regularly become well-known among their peers and gain more prestige over time. Eventually, anything they say is likely to receive positive attention. However, moderators admitted that salience could sometimes become harmful for the group if prominent members dominate discussions and draw all the attention to themselves, while others do not receive attention and feedback to their comments. One moderator said that she would encourage less visible members by liking and commenting on their posts:

When someone famous in the group posts something, she gets loads of likes, whereas someone who's not so good at expressing her ideas and does not have the same status receives no reaction. I try to be on the side of the underdog and comment with something positive like "yeah, that's great" just to show some empathy. (Interviewee 6)

Contrary to traditional media, the success of social media sites relies on users' activity. Even though moderators possess a considerable amount of power over discussion within the group, its members are not completely powerless, and they can influence the course the group takes through their participation. In the interviews, the moderators mentioned a couple of ways how moderated members or users who had been removed would resist their moderation policy. The first is by flagging and reporting content to Facebook's own moderation team. Sometimes, when content receives a number of flags,

it will be removed, even if it is not against Facebook own policy. Flagging content is thus used to bypass local group moderators and question their power (see also Gillespie, 2018). The second way is to create a competing Facebook group that would discuss the same topics and be intended for the same audience, albeit with a different moderation policy that is tailored to addressing the perceived faults of the original group. Among the groups studied, there were some examples of groups that had been created in protest to some other group's moderation policy as their founders had felt they had been treated unfairly.

Eventually, users have the power to keep communities alive by participating or abandoning them, which makes social media groups highly dependent on their membership, and particularly on those who are active contributors and valued by moderators and their peer members. Users can abandon groups if they are not satisfied, and without users creating and updating content, groups will eventually die. This demonstrates how in the context of social media, the power relationship between gatekeepers and the gated remains dynamic and can change.

5. Discussion and Conclusions

As exposure to news, opinions, and political information increasingly occurs through social media, scholars have expressed their concern about its narrowing and polarizing effect on the information that people are exposed to. Research has confirmed social media users' tendency to network with those who have similar opinions, which is identified as a main driving force behind polarization (Boutyline & Willer, 2017; Lewis et al., 2011). These communities of like-minded people are suspected of amplifying individuals' existing beliefs and restricting the free flow of information, which is harmful for the formation of balanced political views, and thus for deliberative democracy. In connection with this scholarly discussion, which often takes place in relation to algorithmic moderation, this study shows how information is filtered in politically motivated grassroots groups by human moderators. Opacity of moderation policy, which has been named as the main problem in the way that platforms conduct moderation (Gillespie, 2018; Roberts, 2016), is also present in moderation done by volunteers. If users are unaware of the filtering that is performed on their behalf, they do not know what information is left out and why, which leads to their participation in public debate being inadequate or even biased.

Social media platforms are major gatekeepers of information as the selection of content is inherent to them. Moderated private groups, such as those presented in this study, provide in many ways a prominent base for the polarization of views, especially if they do not allow dissenting opinions. Controlling access is an effective way to maintain the homogeneity of a group, and hence, moderation may pose a risk that groups eventually develop into echo chambers. Similar to the study

by Kalsnes and Ihlebæk (2021), this study views moderation practices that are concealed from group members as problematic because they are an obstacle for deliberative democracy and personal development. When users are not accepted into a group to begin with, they are moderated and silenced even before they have participated. This study proposes that transparency throughout all decisions and strategies of moderation is important for civic discussion.

The present study has some limitations as the findings rely solely on interviews with moderators. In order to analyze the moderation process as a whole, and better understand power dynamics between gatekeepers and the gated, future research needs to include viewpoints from all stakeholder groups involved, and particularly from those who are the object of moderation.

This study shows that the work of volunteer moderators encompasses a much wider range of activities than simply hiding or removing content, which are named as the main elements of Facebook's approach to moderation (Kalsnes & Ihlebæk, 2021). Controlling access by curating member lists is a major part of moderation in groups on Facebook; however, it has often been left unexplored in prior related studies. Through continuous boundary control, moderators define the group's ideals for those who are inside and outside of group, as well as for themselves. By focusing on their groups' boundaries, the moderators in this study were shown to view outsiders as the biggest threat to the groups and norms. When boundaries are blurred, the existence of the group may be threatened and open to attack. Access to the group is regulated in order to not only maintain group norms and cohesion of views but also to avoid harder forms of moderation. Hard moderation—namely by restricting users' participation or altering it by removing or editing content—occurs in all groups, but because these activities can affect harmony and bring consequences, moderators would rather prevent such incidents by carefully screening potential members. In particular, when the group is in a state that is satisfying for moderators and key members, accepting new members may pose risks.

Volunteer moderators have a challenging task of responding to members' expectations while maintaining the group's main purpose through their everyday moderation tasks. Prior studies have suggested that user-driven communities tend to develop non-democratic structures, so that some users tend to gain more privileges and visibility than others (Keegan & Gergle, 2010; Shaw & Hill, 2014). This study recognizes the existence of "elite members" who have more visibility and power in relation to moderators and get their messages across better than others. Usually, these active members are viewed as beneficial for online communities as these are dependent on their contributions (e.g., Malinen, 2015), but sometimes they can be harmful for the group and its dynamics. A salient member can draw attention to themselves, and so discourage others from contributing.

In the current high-choice social media environment, information transmission has become more direct, but many of the mechanisms through which information flows from producers to users still remain invisible. This study has revealed how volunteer moderators control political discussion in groups on Facebook, and its findings show how they hold a disproportionate amount of power over group members. The relationship between gatekeepers and the gated is thus asymmetrical and unilateral; gatekeepers possess a range of tools for limiting, or even preventing, the participatory opportunities of the gated (see also Dovbysh, 2021). This article has approached gatekeeping in social media as means of control that involves several controlling practices that moderators can use towards group members. Focusing particularly on one of these strategies—boundary control—the findings show how it is used as a discrete but effective way of controlling content. It prevents unwanted users from participating, but at the same time, it is not evident to group members.

Acknowledgments

My warmest thanks to the moderators who participated in this study and shared their experiences. I would also like to thank the anonymous reviewers and the editors of this thematic issue for their useful comments and encouragement in revising this article. This study is part of the author's post-doctoral project NETGATE (2019–2022) funded by Academy of Finland, decision number 321608.

Conflict of Interests

The author declares no conflict of interests.

References

- Barzilai-Nahon, K. (2008). Toward a theory of network gatekeeping: A framework for exploring information control. *Journal of the American Society for Information Science and Technology*, 59(9), 1493–1512.
- Bennett, W. L., & Segerberg, A. (2012). The logic of connective action. *Information, Communication & Society*, 15, 739–768.
- Bessi, A., Petroni, F., Del Vicario, M., Zollo, F., Anagnostopoulos, A., Scala, A., Caldarelli, G., & Quattrociocchi, W. (2016). Homophily and polarization in the age of misinformation. *The European Physical Journal Special Topics*, 225(10), 2047–2059.
- Boutyline, A., & Willer, R. (2017). The social structure of political echo chambers: Variation in ideological homophily in online networks. *Political Psychology*, 38(3), 551–569.
- Choak, C. (2012). Asking questions: Interviews and evaluations. In S. Bradford & F. Cullen (Eds.), *Research and research methods for youth practitioners* (pp. 90–112). Routledge.

- Creswell, J. W. (2013). *Qualitative inquiry and research design: Choosing among five traditions*. SAGE.
- Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarella, G., Stanley, H. E., & Quattrociocchia, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, *113*, 554–559.
- Dovbysh, O. (2021). New gatekeepers in town: How groups in social networking sites influence information flows in Russia's provinces. *Social Media + Society*, *7*(2). <https://doi.org/10.1177/20563051211013253>
- Dutton, W. H. (2009). The fifth estate emerging through the network of networks. *Prometheus*, *27*(1), 1–15.
- Fiesler, C., Feuston, J. L., & Bruckman, A. S. (2015). Understanding copyright law in online creative communities. *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (pp. 116–129). ACM.
- Gachau, J. N. (2016). The role of social media in participatory democracy: A case study of Facebook groups. *Proceedings of the 19th International Conference on Supporting Group Work* (pp. 467–472). ACM.
- Gagliardone, I., Gal, D., Alves, T., & Martinez, G. (2015). *Countering online hate speech*. UNESCO.
- Gillespie, T. (2018). *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.
- Gorwa, R. (2019). What is platform governance? *Information, Communication & Society*, *22*(6), 854–871.
- Gustafsson, N. (2012). The subtle nature of Facebook politics: Swedish social network site users and political participation. *New Media & Society*, *14*(7), 1111–1127.
- Honeycutt, C. (2005). Hazing as a process of boundary maintenance in an online community. *Journal of Computer-Mediated Communication*, *10*(2). <https://doi.org/10.1111/j.1083-6101.2005.tb00240.x>
- Jhaver, S., Appling, D. S., Gilbert, E., & Bruckman, A. (2019). "Did you suspect the post would be removed?" Understanding user reactions to content removals on Reddit. *Proceedings of the ACM on human-computer interaction*, *3*(CSCW), 1–33.
- Kalsnes, B., & Ihlebæk, K. A. (2021). Hiding hate speech: Political moderation on Facebook. *Media, Culture & Society*, *43*(2), 326–342.
- Keegan, B., & Gergle, D. (2010). Egalitarians at the gate: One-sided gatekeeping practices in social media. *Proceedings of the 2010 ACM conference on Computer Supported Cooperative Work* (pp. 131–134). ACM.
- Kushin, M. J., & Kitchener, K. (2009). Getting political on social network sites: Exploring online political discourse on Facebook. *First Monday*, *14*(11). <https://doi.org/10.5210/fm.v14i11.2645>
- Langlois, G., Elmer, G., McKelvey, F., & Devereaux, Z. (2009). Networked publics: The double articulation of code and politics on Facebook. *Canadian Journal of Communication*, *34*(3). <https://doi.org/10.22230/cjc.2009v34n3a2114>
- Lewis, K., Gonzalez, M., & Kaufman, J. (2011). Social selection and peer influence in an online social network. *Proceedings of the National Academy of Sciences*, *109*(1), 68–72.
- Malinen, S. (2015). Understanding user participation in online communities: A systematic literature review of empirical studies. *Computers in Human Behavior*, *46*, 228–238.
- Matias, J. N. (2019). The civic labor of volunteer moderators online. *Social Media + Society*, *5*(2). <https://doi.org/10.1177/2056305119836778>
- Meraz, S., & Papacharissi, Z. (2013). Networked gatekeeping and networked framing on #Egypt. *The International Journal of Press/Politics*, *18*(2), 138–166.
- Myers West, S. (2017). Raging against the machine: Network gatekeeping and collective action on social media platforms. *Media and Communication*, *5*(3), 28–36.
- Pariser, E. (2011). *The filter bubble: What the internet is hiding from you*. Penguin Press.
- Park, N., Kee, K. F., & Valenzuela, S. (2009). Being immersed in social networking environment: Facebook groups, uses and gratifications, and social outcomes. *Cyberpsychology & behavior*, *12*(6), 729–733.
- Pruchniewska, U. (2019). "A group that's just women for women": Feminist affordances of private Facebook groups for professionals. *New media & society*, *21*(6), 1362–1379.
- Roberts, S. T. (2016). *Commercial content moderation: Digital laborers' dirty work*. <https://ir.lib.uwo.ca/commpub/12>
- Ruckenstein, M., & Turunen, L. L. M. (2020). Re-humanizing the platform: Content moderators and the logic of care. *New media & society*, *22*(6), 1026–1042.
- Seering, J., Wang, T., Yoon, J., & Kaufman, G. (2019). Moderator engagement and community development in the age of algorithms. *New Media & Society*, *21*(7), 1417–1443.
- Shaw, A., & Hill, B. M. (2014). Laboratories of oligarchy? How the iron law extends to peer production. *Journal of Communication*, *64*, 215–238.
- Sormanen, N., & Dutton, W. H. (2015). The role of social media in societal change: Cases in Finland of fifth estate activity on Facebook. *Social Media + Society*, *1*(2). <https://doi.org/10.1177/2056305115612782>
- Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, *50*(3), 755–769.
- Terranova, T. (2000). Free labor: Producing culture for the digital economy. *Social Text*, *18*, 33–58.
- Warren, A. M., Sulaiman, A., & Jaafar, N. I. (2014). Facebook: The enabler of online civic engagement for activists. *Computers in Human Behavior*, *32*, 284–289.
- Weber, H. L. (2011). Missed cues: How disputes can socialize virtual newcomers. *Language@ Internet*, *8*(5), 1–18.

Zhu, Q., Skoric, M., & Shen, F. (2017). I shield myself from thee: Selective avoidance on social media dur-

ing political protests. *Political Communication*, 34(1), 112–131.

About the Author



Sanna Malinen is postdoctoral researcher in economic sociology at the University of Turku, Finland. She has studied online communities for over a decade, focusing particularly on users' collective information production and their dynamic roles as information providers and consumers. Her ongoing postdoctoral project examines how information is selected, controlled, and circulated in online public sphere, and how these practices shape power dynamics between gatekeepers and the gated.