The owners of information: Content curation practices of middle-level gatekeepers in political Facebook groups

Sanna Malinen iD
University of Turku, Finland

Abstract

Volunteer moderators play a key role when making judgements about which online content should be accepted and which should be removed. As such, their work fundamentally shapes the digital social and political spheres. Using the data obtained from 15 Facebook group moderator interviews as research data, this study focused on the content curation work by the middle-level gatekeepers of Finnish political discussion groups on Facebook. The findings show that the moderators feel strong ownership of the groups they moderate and of the information such groups provide, and as a result, they strongly shape the groups' discussion and governing policy. Facebook's governing policy for groups is vague, which gives space for group norms and identities to develop. The stakeholder groups (i.e. the platform administration, moderators and users) do not attend to the governance process all together, so negotiations among them are almost non-existent.

Introduction

Discussion and expression of political opinions have increasingly moved online. Citizen participation in social media has an upside: the unmediated and bottom-up nature of social media discussion is expected to expand political participation and thus invigorate

Corresponding author:
Sanna Malinen, Department of Social Research, University of Turku, Turku 20014, Finland.
Email: sanna.malinen@utu.fi

democracy by broadening the range of views and democratising participation (e.g. Benkler, 2006; Hindman, 2009; Tucker et al., 2017). At its prime, online discussion is a free flow of ideas, but people need protection from harmful content and behaviour. Thus, building a safe and attractive online environment requires active moderation. In addition to protecting users, social media moderation is important for maintaining social media's good reputation among users, advertisers and the public at large (Gillespie, 2018: 5; Roberts, 2016). Due to these benefits of social media moderation, research has sought suitable tools and strategies for realising it (Jhaver et al., 2019; Matias and Mou, 2018). However, because they employ content moderation, social media platforms do not only facilitate social activity and bring people together but also control people's behaviour by intervening in information flows and suspending people (Gillespie, 2018).

Mainstream social media platforms like Facebook, Twitter and Reddit have received criticism for lack of clarity in their content moderation practices and policies and in the underlying values of such (Gillespie, 2018; Roberts, 2018). The obscure multi-layer structure of social media platforms' moderation practices and policies involves many actors and systems for content production and moderation. Even though the platforms own the contents at their site, such contents are actually voluntarily created, moderated and shared by the users, and the platforms only turn such contents into commodities by modifying them (Gillespie, 2018: 40–42; Roberts, 2018). As Gillespie (2018: 21) points out, moderation is an essential element of platforms, and because of it, 'everything on a platform is designed and orchestrated', indicating that platforms do not present contents as they are but as individually tailored newsfeeds for each user.

Content moderation is defined as 'governance mechanisms that structure participation in a community to facilitate cooperation and prevent abuse' (Grimmelmann, 2015: 52) and as 'a set of practices used by social media platforms to enforce their guidelines on acceptable content' (Ganesh and Bright, 2020: 8). Leaning on these definitions, this study viewed content moderation as consisting of policies aiming to guide social media behaviour to make it conform to the community guidelines. Content moderation as a tool for social media governance is widely understood to involve the removal of inappropriate contents posted on the platform, but it also involves the removal of the accounts that transmit these contents. Given the scale of major social media platforms, they have developed different moderation strategies to deal with user-generated content. These strategies are editorial review and community flagging by either paid workers or volunteering users and automated detection by software tools (Gillespie, 2018: 77). Algorithmic regulation is a hidden way of selecting what kinds of contents to promote and what kinds of contents to suppress among the users. This form of gatekeeping has been criticised for its impact on political discussion in the 'filter bubble' debate (Pariser, 2011; Tufekci, 2015). As the artificial tools used for detecting violations still lack human sensitivity to cultural values and norms, they sometimes lead to decisions that disrupt group norms (Gillespie, 2018: 169; Myers West, 2017, 2018) As such, the responsibility of social media moderation mainly lies in the community members themselves.

The theoretical body of this study drew from the gatekeeping theory, a framework traditionally used in media studies to describe how information flows through various actors in a media setting (Shoemaker and Vos, 2009; White, 1950). The concept of gatekeeping involves a broad set of practices and strategies and sheds light on the roles and

hierarchies of the actors involved in the information flow (Thorson and Wells, 2016), while the term *moderation* refers mainly to one gatekeeping strategy: removal of certain contents and banning of the people who produced these. Broadly, gatekeeping can be viewed as all the actions that shape information flows. As Gillespie (2018: 5) argues, all kinds of gatekeeping are against the ideals of open Internet and the basic function of social media platforms, which is sharing; as such, platforms moderate reluctantly and behind closed doors.

The term *platform* has been used to denote both the services provided by technology companies and the companies themselves (Gorwa, 2019; Srnicek, 2016). According to Gillespie (2015), platforms typically facilitate access to user-generated contents but do not create them. Today's popular social media platforms play an important role as hosts of user communities (Gillespie, 2018: 17). As noted by Warren and colleagues (2014), Facebook was originally designed for personal networking, but it has recently become an important enabler of civic engagement online. Social media promotes political awareness and communication, which are well exemplified in Facebook groups, the object of this study.

Considering the immense popularity and power of social media platforms in the society these days, it is increasingly important to learn more about the mechanisms behind content selection and control, those not only by media companies themselves but also by the information brokers between the users and the platform: the group owners and moderators. Here, the Facebook group owners and moderators are referred to as 'middle-level gatekeepers' (Hemsley, 2018, 2019). They have a dual role: monitoring the contents to keep their community active and at the same time conforming to Facebook's norms. Looking into their views, this study investigated how gatekeeping is organised within the user communities hosted by Facebook and how it is negotiated between the main stakeholder groups: the platform administration, community owners and members.

Theoretical background

Curation logics

Technological development has urged researchers to update the existing relevant theories and concepts to better describe human activity in the context of digital social spaces. Thorson and Wells (2016) introduced the term *curated flows* to refer to the multiple flows of content present online and offline that are shaped by various actors and affect which information individuals are exposed to. Journalists no longer have a monopoly of the production of news and information, and other sets of curators have become more influential particularly in social media. Thorson and Wells (2016) used the notion of *curation logics* to suggest that these different actors are guided by different incentives and norms that shape their curation decisions on what kinds of content to publish and why. They conceptualised five sets of actors who perform curation: *journalistic, social, personal, strategic* and *algorithmic*. While journalistic curation is guided by professional norms and ethics and aims for balance and objectivity, less is known about the other forms of curation and their underlying logics (Thorson and Wells, 2016). A study on political campaigning on Twitter by Hemsley (2019)

confirmed that different types of actors prefer to share different kinds of content: while journalists are guided by journalistic ethics and act accordingly, there are other politically motivated actors who tend to promote contents that support their political preferences. Hemsley (2018) particularly cites influential social media actors like political blogs that do not aim to be unbiased as their purpose is to promote their political agenda. As their curation logics emphasise political arguments over objectivity, they can become sources of misinformation (Hemsley, 2018).

Most of the prior relevant studies have focused on the gatekeeping processes employed by journalists and professional media; as such, little is known about the curation logics of other social media actors. Social media platforms are criticised for not being as open to their users as they present themselves to be; their governance mechanisms and policies are obscured to maintain an illusion of openness (Gillespie, 2018; Roberts, 2018). Hence, there is a need to learn more about the curation work of actual human gatekeepers and the logics and ideas that guide them in carrying out such work.

This leads to the first research question of this study, as shown below.

RQ1. What is the curation logic that informs the decision making of Facebook group moderators?

Middle-level gatekeepers between the social media platform and the audience

Typical of today's online media environment, gatekeeping processes are networked so that the audience can participate in them, an idea referred to as 'network gatekeeping' (Barzilai-Nahon, 2008). The theory of network gatekeeping recognises the dynamic nature of the relationships between the variant news actors who hold diverse positions and levels of power (Barzilai-Nahon, 2008; Meraz and Papacharissi, 2013). Contrary to the traditional media settings, in the networked online context, non-elites can also become prominent in influencing what is being discussed and how. This has been shown particularly in occasions like mass movements and uprisings driven by social media, where ordinary users have a significant role in raising topics to prominence and elevating others to higher status through active gatekeeping (Meraz and Papacharissi, 2013).

Even though in the networked environment everyone can participate in gatekeeping by selecting information, not everyone has large-enough networks to reach. The network gatekeeping theory suggests that some gatekeepers are more powerful than others. One important factor determining the power positions of gatekeepers is the composition of their social networks, more specifically 'their ability to link networks, communities or clusters together, allowing information to travel far and fast' (Nahon and Hemsley, 2013: 48).

The research on influential social media users acknowledges that influence is gained by holding a visible and central position in the network or by having the authority to support or silence viewpoints (Dubois and Gaffney, 2014; Hemsley, 2019; Huffaker, 2010). In the online context, influentiality is manifested as the ability to trigger feedback, spark conversations and even impact the way a particular topic is talked about (Huffaker, 2010;

Thorson and Wells, 2016). Influential persons thus have the power not only to select information but also to set the topics and shape the flow of conversations. Hemsley (2018, 2019) identified influential social media actors who have the power to shape discussion and information flows and referred to them as 'middle-level gatekeepers'. On Twitter, they are the users who have more followers than the average users and who play an important role as information curators (Hemsley, 2018). They engage in active curation by reviewing information sources both above and below their information class status and transmitting the information to their own networks (Drezner and Farrell, 2008; Hemsley, 2019). Middle-level gatekeepers not only select contents but also reframe, remix and repost them for those who are less well informed; as such, they remind us of the traditional notion of *opinion leaders* (Katz and Lazarsfeld, 1955; Thorson and Wells, 2016). As they are well connected and have a wide reach in networks, their messages spread more widely than do the messages of those without such extent of connectedness and reach (Rogers, 1995), which makes them potential key actors in the spread of information and misinformation (Hemsley, 2018).

Information control as a process reflects the power relations of stakeholders. As Barzilai-Nahon (2008) argues, the relevant research needs to take into account the political power of the stakeholders involved when analysing gatekeeping. This study focused on social media actors who have a central and visible position in their respective networks. The middle-level gatekeepers are influential because they possess the power and tools to shape the discussion and political participation of the ordinary users. In the context of this study, the middle-level gatekeepers are the Facebook group owners and moderators. As they are located between the platform and the group members, they play the role of bridging or connecting these two domains and at the same time balancing the different sets of norms, needs and expectations coming from both sides (Gillespie, 2018; Hemsley, 2018; Matias, 2019). As the Facebook group moderators are often the founders and long-time members of such groups, they enjoy legitimacy and prestige, which makes their position different from those of the ordinary group members (Gillespie, 2018: 126). This study thus asked the question below.

> *RQ2.* How is the status of a middle-level gatekeeper achieved and how do middle-level gatekeepers justify their position?

The human labour behind content moderation

The human labour of social media governance relies either on moderation teams consisting of paid workers who review the contents reported by the users or on volunteer users who report or remove the suspicious contents they come across with. In the first strategy, social media platforms can be held accountable for their policy; whereas, in the second strategy, they have delegated some of their governance power to volunteers (Matias and Mou, 2018). A moderator's position is fundamentally different in these two strategies: in the first strategy, the moderator is an employed anonymous worker of the platform, and in the second strategy, she or he is just a devoted group member, sometimes even its founder. Volunteer moderators are usually personally invested in the community and its

success, have tenure in the community and are valued by its members (Gillespie, 2018: 126). In communities hosted by platforms, volunteer moderators can create and enact their own local policies on acceptable speech, and their work fundamentally shapes the community: its identity and values (Matias and Mou, 2018). Social media platforms give their users a sense of independence and freedom by allowing them to build their own online communities with self-made rules separately from the platform but still depending upon it (Gillespie, 2018: 124–128). This results in the phenomenon of user communities located on the same platform becoming very different from one another in terms of their rules and moderation practices.

One characteristic that distinguishes the social media platforms from the traditional online communities is their structural complexity. Jhaver et al. (2019) denote platforms as distributed moderation systems in which community norms and rules overlap among different communities. For instance, in Reddit, the content moderation is guided by multiple layers of rules, including a user agreement and content policy, a set of established rules defined by the Reddit users (called *Rediquette*) and various community-specific and site-wide rules, in addition to the more implicit behavioural norms (Fiesler et al., 2018; Jhaver et al., 2019). Operating in an ecosystem where rules are derived from multiple sources can result in confusion among the users (Fiesler et al., 2015). A similar distributed content moderation system exists on Facebook, where the users navigate between Facebook's general terms of use and the multiple individually tailored community-created norms and rules.

In the early days of online communities, such communities were formed on the basis of the members' shared interests, with the number of members remaining rather small. Nowadays, however, with the social media platforms having a global user base and an abundance of content, applying management techniques suitable for early online communities does not translate (Gillespie, 2018: 76). The early adopters of Facebook belonged to a rather homogeneous user base of tech-savvy students who shared the same values and norms: those of the Facebook developers (Gillespie, 2018: 117; Kirkpatrick, 2010). As noted by Gillespie (2018: 117–119), it is for this reason that Facebook did not expect content moderation to become a significant problem in the future. Moderation work relies on the active monitoring of contents, which is much simpler in small groups with less content, restricted membership and familiarity among the members (Gillespie, 2018: 76). On Facebook, users encounter the moderation rules and decisions mostly in their day-to-day engagement with the other members of their respective Facebook user communities.

Transparency about moderation rules has been identified as a key element of successful content moderation (Gorwa, 2019; Jhaver et al., 2019; Myers West, 2018). When the social media platform users perceive the online community rules to be clear and understand and accept these, they are more likely to consider the removal of their questionable posts fair and to consider posting again thereafter; lack of clarity in moderation decisions and rules causes frustration in and resistance from the platform users (Jhaver et al., 2019; Myers West, 2017). In reaction to unclear moderation decisions and rules, the platform users start to develop folk theories and suspicions about political biases (Myers West, 2018). There have been many incidents where platform users fought against moderation with visibility, by promoting contents that had been taken down, thus exposing what the

platform was trying to hide (Gillespie, 2018; Myers West, 2018). To address the asymmetrical power relation between the moderators and the moderated, previous studies have proposed involving users in the formulation of the content moderation policy (Matias and Mou, 2018).

According to Matias (2019), volunteer moderating is a form of civic labour in which the volunteers do more than just the work associated with their role; they must also negotiate with the platform operators, the other members of their communities and their fellow moderators regarding how moderation should be approached. Platform moderation is an ongoing negotiation regarding what contents are appropriate and where the limits or boundaries of the group should be drawn in terms of sharing contents on the platform (Gillespie, 2018: 171; Matias, 2019). Matias refers to this daily negotiation as 'boundary work', in which power relations between the stakeholder groups are organised. Drawing from the idea of boundary work as embedded in volunteer moderating, this study asked the question below.

RQ3. How do Facebook group moderators negotiate with the Facebook administration, the other members of the group and their fellow moderators regarding the content-sharing boundaries in the platform?

Data and analysis

The study data were taken from semi-structured interviews with 15 owners and volunteer moderators of active Finnish Facebook groups dedicated to the discussion of political and societal topics. The data were collected from December 2019 to February 2020. The interviewees were selected through the following process. A search was made for groups labelled *political* or *societal*, and then, the names of the groups' moderators and administrators were sought from the groups' public information page. Thereafter, the moderators/administrators were contacted via Facebook Messenger and invited to take part in this study. Eight of them were women and seven were men, and their ages ranged from the early 20s to the 70s. Each of them was interviewed face to face from 1 to 2.5 hours. The interviews were recorded and then transcribed. The interviewees were chosen as the subjects of this study because of their central position in their respective Facebook groups; they had the power to control who would belong to the group, select the contents to be published or posted and shape the course of the online discussion.

The Facebook groups in this study were either local or nationwide in scope, and their sizes varied from 200 to almost 30,000 members. Some of the groups were public and everyone could thus view their discussions, but most of the groups were closed, with their contents visible only to the members. It is noteworthy that most of the interviewees were moderators in more than one Facebook group. As shown in the study by Warren et al. (2014), Facebook is an important communicative tool for gathering and organising information and mobilising people, especially for activists striving to address social issues. The interviewees can also be considered activists; many of them were notable actors in political parties and civic organisations, group

founders and debaters, and for them, social media was just another venue for participation in political and civic activities.

The semi-structured interview strategy allows greater flexibility in the course of a study as there is no fixed range of responses to each question (Given, 2008: 810). It can thus bring out unexpected results and allow the study to take new directions. In semi-structured interviews, the researcher asks the interviewees a series of predetermined but open-ended questions using a planned interview protocol (Given, 2008: 810). The interview protocol that was used in this study followed the main research themes, which were the group's purpose, rules, moderation practices and boundaries and the members' roles. At the end of each interview, the interviewee was encouraged to add comments and bring out new themes if she or he felt that something important was not raised in the interview.

The qualitative data analysis followed thematic analysis, which is a process of identifying patterns of themes within the data. According to Braun and Clarke (2006: 78), 'through its theoretical freedom, thematic analysis provides a flexible and useful research tool, which can potentially provide a rich and detailed, yet complex, account of data'. The analysis first focused on gaining a detailed understanding of the Facebook moderation practices, the moderators' roles and the underlying motivations and values of their moderation work. Following the procedure described by Creswell (2013: 184–187), the text was first aggregated into small categories of information labelled with a descriptive code. Codes were then extracted into broader themes and reviewed several times. Each of the themes were interpreted in terms of their meaning for the research questions, so that, in the final analysis, the main themes that were left contributed to answering them.

Results

Moderators as masters and servants

> I feel that it's my duty as a citizen to participate in the public discussion and express my opinion when something is clearly against my ethical standards. I want to do my share in eliminating hostile and misanthropic tones from public discussion. (Interviewee 12)

Many of the interviewed moderators stated that the purpose of their Facebook group was to increase and spread awareness of the political matters. As most of the interviewees were activists or partisans, they were very open about their political goals and clearly pointed out that they wanted to exert an influence in the society and in the political discussion. In terms of curation logics (Thorson and Wells, 2016), this type of curation can be defined as *strategic* as the groups were openly ideological and were created to promote certain political views or goals, without even trying to be balanced. For example, many small, local Facebook groups were dedicated to either supporting or opposing local political decisions, such as construction projects. However, the interviews revealed an interesting contradiction: even though exerting political influence was the main reason for the creation of all the groups, overly explicit political promotion was an unwanted behaviour in all the groups. This was clearly seen when overly pushy politicians were criticised or even silenced in the groups because the group owners did not want to give too much space and visibility to individual politicians.

The issues of ownership and authority explain the ambivalence of attitudes towards explicit political promotion in groups that are political in nature. The moderators and group owners who were interviewed in this study did not want their groups to be 'hijacked' and become platforms serving the interests of individual politicians. Instead, they emphasised that discussion on political matters should be accessible for all the group members so that everyone could participate and express her or his opinions. They wanted to keep political promotion discrete, and they perceived some individual politicians as too loud and self-centred. As some moderators said, politicians want to make the group discussion revolve around themselves, which does not serve the interests of the group and might even be harmful for the group dynamics if the other members get to feel that their opinions are questioned or attacked.

> There are politicians who use Facebook groups as their mouthpieces or soapboxes; they mainly post their own speeches and recruit people to praise them and diss others in the comments. Some people take offense and leave the thread, so at the end of the day, they just aim to suppress the conversation. (Interviewee 1)

To sum, some of the political Facebook groups are strategic in the sense that they were established to advance the political goals and views of their founders. However, the interviewed moderators of such groups want all the members to act in the best interests of the group, and if individual members are bringing out their personal agenda too aggressively, they are moderated. Facebook groups are a prominent channel for strategic actors for getting their message through, but at least in this study, they are not allowed to bypass the curation by moderators. Particularly, it is perceived as harmful for a group if a more powerful member aims to draw attention from the ordinary members' views and, in this way, make the discussion unbalanced. The moderators want to hold on to their ownership of their groups and do not want to give it to some discussants, particularly if such discussants are powerful and popular politicians.

Another oft-recurring curation strategy relies on the ideal of civilised discussion, in which the opinions are well reasoned and based on facts. This curation style can be referred to as *journalistic* as it aims to encourage deliberation, self-development and learning new ideas through discussion. The moderators act as curators who carefully select valuable pieces of information from various news sources and post these on the site.

> We share fact-based information. I want to give them good sources that they can refer to when they talk about these topics somewhere else. I always try to think of the kinds of content that may be interesting for them, what they want to see in their newsfeeds. I see this group as a news site, a mediator of news, almost like a medium itself. (Interviewee 3)

As quoted above, one interviewed moderator described their moderation team as servants of their audience. She said that like journalists, they aimed to provide their readers with objective and accurate information about their group's themes, and they described their group as a news outlet where people could easily find reliable information. They viewed careful content curation as necessary so that the group members could trust that all the information posted on their group page is of good quality and

has been fact-checked, which contributes to the quality of the discussion. They said they wanted to help their readers pursue their political aspirations and in this way participate in spreading the ideological views of the group.

The most notable difference between the strategic and the journalistic curation styles concerns the objectives of gatekeeping. In the strategic curation logic, the gatekeepers filter the information produced and posted by the members, mainly removing inappropriate contents, whereas in the journalistic logic, the gatekeepers actively search for and collect suitable information from other sources and then post it on their own site. Gatekeeping activities thus slightly differ from one another as strategic curation involves active content removal to protect the group from unwanted contents while journalistic curation focuses on looking for suitable contents for the group.

Moderation is not only about regulating content visibility; the interviews in this study showed that watching and refining the boundaries of a group is a significant part of moderation work. Moderators reported that they prevent trolling and unwanted behaviour by carefully screening membership applicants and devote much time and effort to creating guidelines and rules for detecting individuals who may be harmful for the group. An applicant's suitability to the group is judged based on her or his friendships, profile pictures, other group memberships, liked Facebook pages and so on. By controlling the access to the group, moderators define who is a suitable member and likely to meet their expectations. Member selection is thus an important content curation strategy for moderators; it is both *personal* curation as it relies on moderators' personal preferences and *social* curation because only the members have the permission to post contents to the group.

Boundary control and recognition of potential troublemakers are extremely important for the success of closed groups. Moderators need to maintain the members' trust in one another to encourage them to discuss and share. This type of protective curation strategy aims to build a safe space for the members and to maintain a positive and encouraging environment particularly for those who are in a marginalised or vulnerable position. There were several stories of how groups were harmed by a member who leaked out contents and published screen caps at other discussion boards with the intention of degrading the group and its members. Leaks have a detrimental effect on trust: when the members are afraid of sharing their personal opinions or even joking because their words can be taken out of context and used to bully them, the group may eventually die. This type of public ridiculing is especially targeted at persons with a more public profile; making fun of activists can thus be interpreted as a way to silence them. In Facebook groups discussing polarising topics, a great deal of moderation is about keeping the group's borders closed from outsiders and protecting the members from attacks coming from their perceived enemies. Particularly concerning sensitive discussion topics, the interviewed moderators took it upon themselves to make sure that their respective groups would remain safe spaces in which the members would feel comfortable sharing their opinions with the other members, without fear of being attacked or ridiculed.

> It became a huge problem in the group. People made fake profiles and came into the group just to find juicy tidbits to be leaked. In particular, those who made public appearances and who regularly talked about these things in public were attacked. It wasn't even about any personal

or sensitive stuff; someone just took some comments out of context, reframed them and said, 'Hey, look at this idiot!' (Interviewee 6)

As noted by Thorson and Wells (2016), in today's complex online environment, the same content is circulated by various sets of curators and within different sets of curation logics. An interesting question is *Are there any underlying hierarchies between these curating actors so that some actors and their logics are preferred over others?* The middle-level gatekeepers in this study present the most visible and active part of Facebook's content curation system while the other types of curators remain hidden. The platform's multi-level governance system provides different sets of selection tools to different stakeholders. Group moderators' affordances for doing selection are broader than those of average users are, and they can exert all forms of curation, except algorithmic curation. To answer the RQ1, this study identifies a variety of curation logics among moderators and, focusing on a relatively small sample of political Facebook groups, highlights the role played by moderators' personal views and goals in their curation decisions. Sets of logics thus depend heavily on the group owners; many of those who were interviewed in this study viewed the group that they were moderating as a strategic tool for spreading their political objectives while some used the group for learning new things or exchanging ideas. What was common for all the moderators was that they emphasised the role of active moderation in achieving the group's goal.

Negotiation or dictation?

Many of the interviewees in this study described themselves as professional moderators who have had moderating experience in other Facebook groups. They participated actively in the creation of the group's policy and wanted to develop a more general set of rules to be applied in different Facebook groups. They thus called for universal and standardised rules instead of the current locally varied rules and moderation practices. However, establishing and maintaining a professional role required their withdrawal from discussions and their avoidance of using an aggressive tone. Many of them said that they no longer participated in discussions but resigned so they could follow the discussions as bystanders.

> I have toned down my behaviour and have become much more moderate. Earlier, I was a fiery discussant, but because a moderator has a position of power in the group, I think moderation has a lot to do with power. (Interviewee 10)

Becoming an expert moderator requires distinguishing oneself from the other group members, and building a strong moderation team is one way of doing that. In teams, they collaboratively reviewed contents and negotiated decisions with their fellow moderators particularly in relation to difficult moderation tasks. However, the existence of a professional moderation team working independently of the group members leads to the isolation of the moderation work from the members as decisions are made without the members' participation. The interviewed moderators explained that separating the roles of the moderators and members contributes to the objectivity

and fairness of moderation work. Occasionally, however, the members want to take part in the moderation and have a say on what is being discussed and how. The interviews revealed that tensions arise particularly when the members want to shape and determine the boundaries of the discussion. When the interviewees were asked how the members could participate in the moderation work, most of them made it clear that they did not want the members to take part in the moderation work other than by reporting inappropriate contents, which is helpful for the moderators as they cannot review all the contents posted. Some moderators condemned the idea of peer moderation: the members guiding one another in discussions.

> We want the moderation team to be the one to judge what is allowed and what is not. Sometimes it really annoys me when the ordinary members start to instruct the others because the situation then escalates until it's no longer about the topic but already about the members' individual communication styles, and everything becomes personal, like 'Carl, you can't say that', and then Carl responds, 'No, Peter! Look at what you did here!' Then there are a hundred comments without any point at all. (Interviewee 1)

As the above quotation shows, many of the volunteer moderators who were interviewed considered themselves experts in defining what is an acceptable content or behaviour. They had developed a strong professional identity but did not feel accountable to the members. They often did not acknowledge criticism coming from the group members. In addition, as in the previous studies (Jhaver et al., 2019; Myers West, 2017), the users' reaction to moderation was usually negative, and those whose posts were removed felt that they were treated unfairly. The moderators explained their moderation-related decisions to users when they were asked about these, but they did not change their decisions. According to them, there is no point in arguing with the users. Some previous studies have paid attention to the asymmetrical power relations between the users and moderators of social media platforms and described the users' collective action to resist the moderation decisions they considered unfair (Gillespie, 2018; Myers West, 2017). The interviews revealed that the members of Facebook groups can comment on and criticise the group's moderation policy, but they do not have much power to affect the moderation-related decisions, and if they do not accept moderation, they usually end up leaving the group.

The Facebook group moderators who were interviewed in this study felt a strong sense of ownership of the group that they were moderating and had a clear vision of how it should be governed. One explanation of this strong sense of ownership is how most of the Facebook groups were born. They were founded to meet the goals or needs of the founder, and very often, they were created as a reaction to some other existing group, which the founder considered badly moderated. In the interviews, one oft-recurring story was how a person was suspended from some other group because of her or his views, and how she or he perceived such group's moderation policy as biased or unfair in many ways. In response, the person decided to create a new group that was clearly different from the other one in terms of purpose and rules.

> This moderator is a well-known leftist whom I don't empathise with that much. He once banned me for a while, and when this group started to support political views I do not approve of, I

could not bear it anymore and thus created my own group, which was clearly more conservative. (Interviewee 7)

Particularly in the bigger groups, the moderators had built a moderation team, in which the day-to-day moderation decisions were made. The team usually consisted of group founders and long-time members whose opinions and conversation styles were well known. Usually, the team invited reliable and active discussants as moderators. In some groups, the teams were created carefully considering each moderator's political preferences so that the invited moderators would represent as diverse political backgrounds as possible. By doing this, they mitigated the *echo chamber* effect and aimed to make moderation decisions that were unbiased and fair. Below is a direct quote from the moderator of a group with a broad spectrum of political views.

We divided the work so that if a green-left-leaning person is losing it, I as a green-leftist will handle the situation. If a right-wing supporter gets out of control, I cannot moderate that because we don't want moderation to appear politically motivated, and obviously they would blame me for being biased. Before we mute or kick someone out of the group, though, the majority of the team has to agree with the decision. (Interviewee 14)

Peer moderators play a key role in judging whether to accept members and whether to remove a post or ban a person. The interviewed moderators relied on the opinions of the other peer moderators, and the disagreements were very few. They wanted their decisions to be uniform, which was perceived as an indicator of the group's functionality. In the groups with only one moderator, the moderation system was obviously simpler, and the questions of balance and fairness were not considered in the same way as the moderation policy was: based only on one person's judgement. One of the moderators, partly jokingly, referred to himself as a dictator, meaning he does not negotiate about content removals or bans. However, he spent much time explaining to irate members why they were punished and which rule they violated. The interviews thus showed that the strong moderation teams with clear rules and that were striving for political objectivity did not need to explain or justify their decisions as much as the dictator-type moderators did.

In the interviews, moderators indicated that they had developed a strong professional identity, and over time, they had refined their working methods to be considered as objective and considerate experts by other group members. Many moderators said that they wanted to distinguish themselves from the other group members and avoided participation in the group discussions to maintain their professional role. Instead, moderation lines were drawn within moderator teams in which they discussed their decisions with fellow moderators. To answer the RQ2, distancing oneself from the actual discussion and making decisions within the moderator team were techniques that contributed to their expertise and objectivity and justified their middle-level gatekeeper position in the groups.

Faceless Facebook

Volunteer moderators of this study quite uniformly stated that they do not communicate with Facebook representatives. The moderators pointed out that the Facebook

administration is faceless and that its moderation policy is obscure. They said that even though Facebook regularly updates its terms and policies, they perceive the platform itself as detached from the group; thus, they did not even think about consistency with the platform policy when they were creating their own moderation rules. Encounters with the Facebook moderation team were not common among the interviewees, but some of them reported having seen occasional content removals by Facebook in their group without any explanation. The moderators considered some of Facebook's removals inconsistent and having unclear bases, but they did not think that these were worth arguing with the Facebook administration over. For example, Facebook removed rather harmless contents (e.g. caricatures) from their group page, but on the other hand, when some of the interviewees reported offensive contents, Facebook did not respond.

> I reported to Facebook one message that encouraged violence, but I was told that it was not against their community norms. That was strange; I have no idea how Facebook's content moderation works. (Interviewee 5)

Some of the interviewed moderators suspect that Facebook's moderation is based on automatic detection, which may explain why Facebook tends to react to pictures rather than to written text. Particularly, if a text is written in Finnish, it is most likely left unattended, whereas pictures presenting nudity, violence or controversial symbols are quickly removed.

Two of the interviewed moderators had encountered the hardest form of Facebook moderation: removal of the whole group. They suspected that the reason for Facebook's removal of their group from the platform was that their group posted a link to a content from a well-known countermedia site, and Facebook defined it as fake news. Facebook first sent a warning about the removal; soon after, it removed the group from the platform. As one of the interviewed moderators said, Facebook does not negotiate with the moderators or explain its decisions.

> We went to see the Facebook norms after the group was removed, and it says that sharing false information is not allowed. That may be the reason for our group's removal. As such, even if the story itself does not consist of false information, the Facebook rule says that you cannot publish it because of its source. It's the same with pictures; if you post a controversial picture and want to talk about it in a critical manner, Facebook will remove it even though your intention is not offensive. (Interviewee 11)

In one group, where the users frequently reported offensive contents to Facebook, the moderators felt uneasy, feeling that they were under tightened surveillance. Facebook had given them several warnings and had frequently removed contents from their group. The interviewed moderator suspects that the group is likely to disappear soon due to only a minor violation.

Volunteer moderators have the challenging task of trying to meet both Facebook's and the group members' expectations while maintaining the group's purpose in their everyday moderation work. The moderators in this study said that they wanted their respective groups to reflect their own values and political views, and most of the moderators had invested much time and effort in the creation of the rules and policies of their group.

However, Facebook groups are not as independent from the platform's guidelines as their moderators and members may think. As Gillespie (2018: 127) noted, the group administration can be easily bypassed using the platform-wide flagging mechanism, in which the users report contents directly to Facebook. Directly reporting contents to Facebook can be a problem particularly for political groups. In controversial discussions, a group's political enemies can flag and report group contents issuing unwanted opinions and can even make the whole group disappear to advance their political goals in the name of moderation. Some of the interviewed moderators suspect that the people who are against their group and its ideology are using flagging as a means to eliminate political views contrary to theirs.

> Sometimes Facebook notifies us that it has removed a certain post from our group page because it's against Facebook's norms. The posts that are removed are usually politically provocative, but they are really not against Facebook's rules. We can clearly see when outsiders have reported such posts to Facebook, and we know that they're playing a dirty game. It's not like Facebook is some court of justice, though; it's quite arbitrary when it bans and takes down contents. (Interviewee 12)

The community norms and moderation policies of the social media actor Facebook are vague to the users and remain such until someone violates them. The interviewed moderators did not show much interest in Facebook's own moderation policy nor expected Facebook to intervene in how they were running their groups. Hence, Facebook has succeeded in giving them the illusion of freedom in organising and nurturing their own communities (Gillespie, 2018: 126). However, when a group's local moderation policy collides with Facebook's, the group's policy will be overridden. As the moderators claimed, when Facebook intervenes in problems, the local group administration is completely forgotten: Facebook does not negotiate or explain its decisions to the group moderators and members.

Contrary to what was suggested by prior research (Matias, 2019), not much collaboration among stakeholder groups occurs in moderation processes. With regard to the RQ3, the interviewees of this study indicated that they negotiated moderation lines within the moderation team and appreciated the opinions of the moderator colleagues, but there is very little negotiation with the members and no communication at all between the platform government and the moderators.

Conclusion

This study has shed light on the curation logics that guide content selection in political Facebook groups. The findings highlight the power of the middle-level gatekeepers, showing that their work is about not only selecting or removing contents but also controlling who will have access to contents. For moderators, selecting who can join the group is an important way of doing curation, as it is about choosing who has the permission to produce contents to the group. The term 'boundary work' (Matias, 2019) is thus an appropriate term to use for content moderation, indicating that a big part of it appears to involve drawing the boundaries of the group and deciding who can join it and who cannot. This study showed that content moderation in Facebook groups is

mainly protective as its goal is to protect the group members from hostile behaviour and attacks from external agents.

This study confirmed that Facebook's governance policy remains a *black box*: unclear to both the Facebook group members and moderators. Perhaps due to Facebook's invisibility, the group moderators who were interviewed in this study said that they felt a strong ownership of the group that they were administering and did not expect any interference from Facebook. The moderators viewed themselves as professional community managers and justified their decisions with the expertise that they had developed over time. It is noteworthy that the decisions about content moderation are made within the moderation team and that the users are expected to stick to their role and to participate in content moderation only by flagging and reporting offensive contents to the moderators. Unlike the previous study by Matias (2019), which conceptualised volunteer moderation as a constant negotiation among the stakeholders, this study drew a very different portrait as according to the moderators, not much negotiation takes place among the social media platform, the middle-level gatekeepers and the users in terms of the moderation processes to be implemented. All the stakeholder groups involved in moderation appear to be distant to each other. This is partly due to the dispersed moderation system in which the platform representatives remain unknown to others and also because moderators wanted to distance themselves from the actual discussions and make the moderation decisions without members.

Previous work on moderation has highlighted the importance of transparency for successful moderation. In this study, even though transparency was lacking from the platform's part, moderators did not want its involvement to become more visible in groups. The moderators indicated that they do not want Facebook to improve its moderation in terms of becoming more active in defining and implementing its own governing policy in groups. Even though they considered Facebook as invisible and distant, they wanted to maintain their freedom in creating group policy. Thus, when developing platform governance, it is important to keep in mind that strong sense of ownership felt by volunteer moderators may be crucial to their willingness to create and manage user groups, which they often described as mentally demanding and time-consuming work.

The power relations among the social media platform, group moderators and users become visible when someone violates the rules. Like a previous study by Myers West (2017), this study confirmed the existence of asymmetrical power relations among the stakeholders. The findings show that the middle-level gatekeepers strongly shape the discussion in the Facebook groups and that the members have little influence on such groups' governance. In view of the powerful leadership, it can be said that the participation structure of the Facebook groups is inherently hierarchical rather than democratic as the group leaders can assert their authority without clearly defined institutions or boundaries (Shaw and Hill, 2014). When 'invisible' Facebook removes groups or contents, the moderators and members have no choice but acquiesce. The interviewed moderators unanimously pointed out that Facebook neither negotiates nor explains its decisions and that if there is a disagreement, Facebook always wins. Facebook is currently hosting a whole ecosystem of local groups created to fulfil their owners' personal goals and meet their needs. Despite all the efforts and expertise of the members and owners of such groups, Facebook always has the last word in terms of the content moderation policy.

Funding

ORCID iD

Sanna Malinen (iD) https://orcid.org/0000-0003-4644-0004

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.

Benkler Y (2006) *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. New Haven, CT: Yale University Press.

Braun V and Clarke V (2006) Using thematic analysis in psychology. *Qualitative Research in Psychology* 3(2): 77–101.

Creswell JW (2013) *Qualitative Inquiry and Research Design: Choosing Among Five Traditions*. Thousand Oaks, CA: SAGE.

Drezner D and Farrell H (2008) The power and politics of blogs. *Public Choice* 134(1–2): 15–30.

Dubois E and Gaffney D (2014) The multiple facets of influence: identifying political influentials and opinion leaders on Twitter. *American Behavioral Scientist* 58(10): 1260–1277.

Fiesler C, Feuston JL and Bruckman AS (2015) Understanding copyright law in online creative communities. In: *CSCW'15: Proceedings of the 18Th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 14–18 March 2015, Vancouver, BC, pp. 116–129. New York: Association for Computing Machinery.

Fiesler C, Jiang JA, McCann J, et al. (2018) Reddit Rules! Characterizing an Ecosystem of Governance. In: *Proceedings of the International AAAI Conference on Web and Social Media*, Palo Alto, CA, 25–28 June 2018. Palo Alto: AAAI Press.

Ganesh B and Bright J (2020) Countering extremists on social media: challenges for strategic communication and content moderation. *Policy & Internet* 12(1): 6–19.

Gillespie T (2015) Platforms intervene. *Social Media + Society* 1(1): 1–2.

Gillespie T (2018) *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. New Haven, CT: Yale University Press.

Given LM (ed.) (2008) *The SAGE Encyclopedia of Qualitative Research Methods* (Vols. 1–0). Thousand Oaks, CA: SAGE.

Gorwa R (2019) What is platform governance? *Information, Communication & Society* 22(6): 854–871.

Grimmelmann J (2015) The virtues of moderation. *Yale Journal of Law & Technology* 17(1): 42.

Hemsley J (2018) The role of middle-level gatekeepers in the propagation and longevity of misinformation. In: Southwell BG, Thorson EA and Sheble L (eds) *Misinformation and Mass Audiences*. Austin, TX: University of Texas Press, pp. 263–273.

Hemsley J (2019) Followers retweet! The influence of middle-level gatekeepers on the spread of political information on Twitter. *Policy & Internet* 11(3): 280–304.

Hindman M (2009) *The Myth of Digital Democracy*. Princeton, NJ: Princeton University Press.

Huffaker D (2010) Dimensions of leadership and social influence in online communities. *Human Communication Research* 36(4): 593–617.

Jhaver S, Appling DS, Gilbert E, et al. (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: 1–33.

Katz E and Lazarsfeld PF (1955) *Personal Influence: The Part Played by People in the Flow of Mass Communication*. Glencoe: Free Press.

Kirkpatrick D (2010) *The Facebook Effect: The Inside Story of the Company That Is Connecting the World*. New York: Simon and Schuster.

Matias JN (2019) The civic labor of volunteer moderators online. *Social Media + Society* 5(2): 2056305119836778.

Matias JN and Mou M (2018) CivilServant: community-led experiments in platform governance. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Paper 9), pp. 1–13. New York, NY: ACM. DOI: 10.1145/3173574.3173583

Meraz S and 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.

Myers West S (2018) Censored, suspended, shadowbanned: user interpretations of content moderation on social media platforms. *New Media & Society* 20(11): 4366–4383.

Nahon K and Hemsley J (2013) *Going Viral*. Cambridge: Polity.

Pariser E (2011) *The Filter Bubble: What the Internet Is Hiding From You*. New York: Penguin Books.

Roberts ST (2016) Commercial Content Moderation: Digital Laborers' Dirty Work. In: Noble SU and Tynes B (eds) *Intersectional Internet: Race, Sex, Class and Culture Online*. New York: Peter Lang, pp. 147–160.

Roberts ST (2018) Digital detritus: 'Error' and the logic of opacity in social media content moderation. *First Monday* 23. 3–5 March 2018. https://firstmonday.org/ojs/index.php/fm/article/download/8283/6649

Rogers EM (1995) *Diffusion of Innovations*. New York: Free Press.

Shaw A and Hill BM (2014) Laboratories of oligarchy? How the iron law extends to peer production. *Journal of Communication* 64: 215–238.

Shoemaker PJ and Vos TP (2009) *Gatekeeping Theory*. New York: Routledge.

Srnicek N (2016) *Platform Capitalism*. Cambridge: Polity Press.

Thorson K and Wells C (2016) Curated flows: a framework for mapping media exposure in the digital age. *Communication Theory* 26(3): 309–328.

Tucker JA, Theocharis Y, Roberts ME, et al. (2017) From liberation to turmoil: social media and democracy. *Journal of Democracy* 28(4): 46–59.

Tufekci Z (2015) Algorithmic harms beyond Facebook and Google: emergent challenges of computational agency. *Colorado Technology Law Journal* 13(2): 203–218.

Warren AM, Sulaiman A and Jaafar NI (2014) Facebook: the enabler of online civic engagement for activists. *Computers in Human Behavior* 32: 284–289.

White DM (1950) The 'gate keeper': a case study in the selection of news. *Journalism Quarterly* 27(4): 383–391.

Author biography

Sanna Malinen is a Ph.D. in information sciences and Postdoctoral Researcher in Economic Sociology at the University of Turku, Finland. She has studied user behaviour in online communities for over a decade, focusing particularly on people's collective information production practices and their dynamic roles as information providers and consumers. Her ongoing project examines power dynamics between gatekeepers and the gated in online public spheres.