

Ahmad,Sana; Krzywdzinski, Martin

**Book Part — Published Version**

## Moderating in Obscurity: How Indian Content Moderators Work in Global Content Moderation Value Chains

**Provided in Cooperation with:**

WZB Berlin Social Science Center

*Suggested Citation:* Ahmad,Sana; Krzywdzinski, Martin (2022) : Moderating in Obscurity: How Indian Content Moderators Work in Global Content Moderation Value Chains, In: Graham, Mark Ferrari, Fabian (Ed.): Digital Work in the Planetary Market, ISBN 978-0-262-36982-4, The MIT Press, International Development Research Centre, Cambridge, MA, Ottawa, pp. 77-95, <https://doi.org/10.7551/mitpress/13835.003.0008>

This Version is available at:

<https://hdl.handle.net/10419/256896>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



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

## 5 Moderating in Obscurity: How Indian Content Moderators Work in Global Content Moderation Value Chains

Sana Ahmad and Martin Krzywdzinski

Cairncross's (2001) observation that "companies will locate any screen-based activity anywhere on earth, wherever they can find the best bargain of skills and productivity" has often been cited to illustrate the reach of offshoring and international outsourcing. Content moderation is an important part of this global "outsourcing complex" (Peck 2017) that exemplifies its dynamic structures in terms of creation of new digitally mediated economic geographies and continual changes to the contours of the planetary labor market.

As Roberts (2019) and Gillespie (2018a) emphasize, content moderation is a core process for maintaining the social media platforms that have become a central element of contemporary social exchange and global public communication. Despite the enormous social and political importance of the practice, relatively little is known about content moderation. The available bits of public information about it are hidden within the self-declared statements and transparency reports from the social media firms. Content moderation has often been understood in the public discourse as an automated task; the importance of human content moderation has been highlighted by scholars and journalists only recently.

The increasing interest in content moderation has been driven by scholars arguing that social media are part of the public sphere, and therefore the rules governing them should not be set by firms following narrow economic interests (Klonick 2017; Gillespie 2018b). However, the focus of this chapter is not consider in detail whether the onus of user content management should be put on social media firms (Shepherd et al. 2015). The focus is rather to take a step further to untangle cryptic clues about actual outsourcing practices.

Trying to understand why social media firms outsource content moderation service work, to whom they outsource it, and how they find their outsourced contractors is as complicated as unpacking the technological tools and software that are used for moderating content. In this chapter, we analyze the relations between social media firms and their suppliers in India based on the global value chain (GVC) approach (Gereffi,

Humphrey, and Sturgeon 2005; Ponte and Sturgeon 2014). Following an analysis of content moderation value chains, we consider the content moderation labor process and analyze the attendant working conditions.

This chapter focuses on India, which—along with the Philippines—has become a core destination of content moderation outsourcing, according to various investigative scholarly works (Ahmad 2019) and articles in the media.<sup>1</sup> These are the main questions guiding the analysis:

1. What are the value chain configurations through which global social media firms and suppliers in India coordinate content moderation services?
2. How is the labor process of content moderation organized and controlled within these value chain configurations?

First, we define content moderation and place its importance for social media firms. We then outline our motivations for using GVC theory and the labor process approach to analyze the outsourcing of content moderation to India. After providing an overview of how the empirical research was conducted, we develop our analysis of content moderation value chain configurations, with a focus on economic upgrading possibilities for the suppliers in India. We then describe the content moderation labor process with special attention to the working conditions and mobilities of Indian content moderators. In the conclusion, we revisit our main arguments and recommend the role of public and private policies in regulating outsourced content moderation practices.

### Defining Content Moderation

According to Sarah T. Roberts (2017, 1), “Content moderation is the organized practice of screening user-generated content posted to Internet sites, social media, and other online outlets, in order to determine the appropriateness of the content for a given site, locality, or jurisdiction.” The global content moderation market is segmented into different moderation types, primarily *proactive* moderation (before the content is published on the site) and *reactive* moderation (after the content is published on the site) (Grimmelmann 2015). Moderating content has transitioned from open and voluntary moderation of text-based social communities (Usenet groups, Wikipedia, etc.), to the deployment by media firms of word filter technologies along with human skills for comment moderation, to contemporary large-scale moderation practices that use professional moderators and basic algorithms (Roberts 2017).

Roberts (2019) notes that content moderation is a crucial aspect of protecting corporate identity, maintaining the operational laws of platforms, and sustaining as well as

driving an increase in the number of users and their activity on the platforms. Gillespie (2018a) goes a step further in stating that moderation is an “essential, constitutional and definitional” aspect of what platforms do, adding that this content moderation work is hard because it is “resource intensive and relentless.”

On the one hand, content moderation is a tremendously sensitive practice that lies at the core of the activities of social media platforms like Facebook and YouTube. On the other hand, as Sam Levin writes in the *Guardian*, “Silicon Valley has stuck to its foundational belief that tech firms are ultimately not accountable for the content on their platforms.”<sup>2</sup> Indeed, the Silicon Valley-based social media giants have since their inception enjoyed nonliability through Section 230 of the Communications Decency Act in the United States. This regulation has offered these platforms and other websites a “safe harbor” from liability because the third-party postings or content generated by users on these platforms is not their responsibility. For some, Section 230 has also meant a “marketplace orientation” where these platforms can take advantage of the free speech paradigm because “suppression of speech can be anathema to the marketplace theory” (Medeiros 2017, 2).

However, international scrutiny of social media platforms’ moderation policies has increased as a result of many legal regulations, notably Germany’s Network Enforcement Act (*Netzwerkdurchsetzungsgesetz*), India’s intermediary liability regulations, and Iran’s demand that the messaging platform Telegram relocate its distribution networks in the country. In a similar vein, a number of governments, including those of India and Indonesia, have increasingly pressured the Chinese social media firm ByteDance to set up domestic content moderation units in compliance with local regulations. Analyst Kalev Leetaru has gone a step further and likened social media platforms to “opaque black boxes into which we have absolutely no insight or voice.”<sup>3</sup> Such statements, however, reduce users, content moderators, and other engaged participants to powerless positions while ignoring the economic incentives, audience interests, and wider political movements that influence these platforms (Gillespie 2018a).

## Global Value Chains and the Labor Process Approach

### Adopting a Global Value Chain Approach

The outsourcing policies of social media firms have created global content moderation value chains as a part of the planetary labor market (Graham and Anwar 2019). In these value chains, we can distinguish three major actor types: social media firms, multinational enterprises (MNEs), and content moderation suppliers (figure 5.1).

The lead actors are the social media firms (e.g., Facebook, Google)—mainly based in the US but also in Europe and China—that control social media platforms and the



**Figure 5.1**

Idealized content moderation value chain.

*Source:* Authors.

standards and software infrastructures that guide the work of content moderators. In some cases, these firms cooperate with MNEs that provide IT solutions and business process outsourcing (BPO) services. MNEs (e.g., Genpact, Accenture) usually have their headquarters in the Global North and subsidiaries in the Global South. There are also, however, Indian firms providing BPO services related to content moderation (e.g., Foiwe Info Global Solutions). The MNEs provide content moderation services themselves and also outsource it to domestic content moderation suppliers in India or other countries like the Philippines. In some cases, social media firms engage directly with Indian (or Philippine) content moderation suppliers.

India is one of the main destinations of content moderation outsourcing, accounting for over a tenth of moderation workers worldwide.<sup>4</sup> In addition to this outsourcing of content moderation for markets in the Global North, there is an expanding content moderation market in India itself. Indeed, a greater expansion of the mobile phone user base in rural and non-English-speaking regions of India (Tenhunen 2018), together with cheap mobile Internet packages, has sparked growth in content in vernacular languages.

Content moderation GVCs are characterized by high power asymmetries between the lead social media firms in the Global North and the content moderation suppliers in the Global South; these asymmetries influence employment and working conditions. A recent contribution to GVC theory (Ponte and Sturgeon 2014) argues that these power asymmetries are based on a number of factors. From the supplier perspective, one factor is related to technological and organizational capabilities. Most content moderation suppliers do not have technological capabilities and depend completely on software infrastructures provided by the lead firms. But even in cases of more capable suppliers, the monopolistic or oligopolistic structure of lead firms, combined with the low costs for the firms of switching between suppliers (in the case of highly codified and standardized transactions), leads to huge power asymmetries between lead firms and suppliers (Ponte and Sturgeon 2014).

With content moderation practices treated as industrial secrets, there is a lack of scholarship about the structure of content moderation value chains. Social media clients declare several reasons for this secrecy, such as protecting the identities of workers (Gillespie 2018b), preventing users from “gaming the rules” when posting illicit content on social media platforms (Roberts 2016), and safeguarding their proprietary technology. We can, however, build on existing research about the position of Indian firms in value chains shaped by IT BPO, which includes call center firms and others (Noronha and D’Cruz 2020).<sup>5</sup>

The GVC literature has developed a complex classification of value chain structures (Gereffi, Humphrey, and Sturgeon 2005). In the case of BPO-based value chains, two types are of particular relevance. Value chains with low supplier capabilities, a high codification of transactions, and strong direct control of suppliers by lead firms are characterized as “captive.” They do not leave space for the upgrading of the suppliers’ capabilities and are often associated with low-road approaches regarding wages and working conditions (Schrage and Gilbert 2019, 206). Value chains with a high codification of transactions but rather arm’s-length relations between lead firms and suppliers, low direct control, and low switching costs are characterized as “market-based.”<sup>6</sup> Some authors (Lakhani, Kuruvilla, and Avgar 2013, 453) argue that market-based value chains are associated with employment systems relying on “moderately skilled workers” and providing better working conditions than captive ones. Ponte and Sturgeon (2014) emphasize, however, that low switching costs between suppliers can constitute the basis for high power asymmetries between lead firms and suppliers, leading to high price competition, low wages, and bad working conditions.

### **Working Conditions in the Indian IT Sector: A Labor Process Perspective**

A key argument of our chapter is that the opacity of the content moderation industry affects the working conditions of content moderators. There is a lack of systematic studies on the topic that draw on established sociological concepts such as labor process theory. Only media reports about the great psychological stress caused by this form of work have drawn public attention to this area of employment.<sup>7</sup>

We do, however, have a longer history of research into the broader Indian IT BPO sector. Labor process analyses of India’s call center firms have provided useful insights into the work organization and management strategies of these firms and the country’s position in the (then) new international division of labor (Batt et al. 2005; Noronha and D’Cruz 2006). In their discussion on the characteristics of labor processes in call center firms, Batt et al. (2005) emphasized the vulnerability of the workforce due to the subordinate position of Indian firms in GVCs. As Taylor and Bain (2005) have pointed

out, standardized, simple, and tightly scripted tasks have been outsourced to India, and case studies have described tight technological control of the labor process leading to increased work pressure and stress (Taylor and Bain 2005; Noronha and D'Cruz 2006).

Shehzad Nadeem's (2011) ethnographic study of the Indian call center industry looks at the proletarianization of white-collar workers who are engaged in rote customer service work. The 24-hour work cycle encourages long working hours and night shifts, leading to workers' estrangement from social ties and the normal rhythms of life. Nadeem's study not only captures workers' frustrations and their exploitation but also tries to grapple with the dialectic of outsourcing in India. For him, the "concrete realities of a particular place" in India have been transformed into a space of capital accumulation (Nadeem 2011). The creation of spaces such as India's special economic zones is designed to obfuscate the workers both from their counterparts in outsourcing nations and from consumers.

At the same time, existing studies on the Indian IT sector confirm a central argument of labor process analysis: that even highly standardized and technically controlled work processes still require the agency of the workers—that is, their problem-solving ability and participation in coping with the innumerable situations in which standards and technical controls are insufficient (Smith and Thompson 1998; Thompson 2003). Research on call centers has particularly highlighted the emotional work of Indian call center agents in managing stress and unforeseen difficulties in communicating with customers (D'Cruz and Noronha 2008; Remesh 2008). On this basis, Indian workers are trying to work out a path toward professionalization (D'Cruz and Noronha 2013), but this remains difficult given the market-based and captive value chains in the Indian IT sector (Taylor and Bain 2005; Upadhyaya 2010). We expect to encounter similar developments in Indian content moderation companies.

## **Data and Methods**

This study is based on 35 interviews with content moderators, representatives from suppliers located in India, domestic social media firms, Indian trade unions, and civil society organizations.<sup>8</sup> In order to understand the position of Indian firms within GVCs, we started our empirical analysis with interviews with management representatives from social media firms and Indian content moderation suppliers. We conducted interviews with two domestic social media firms. DSM-01 is a medium-sized firm (50–249 employees) that currently handles over 150 million user accounts, and DSM-02 is a small firm (10–49 employees) that went out of business in 2019. In addition, we elicited data through an email exchange with the policy communications manager of

an American conglomerate (GSM-01), which owns the biggest social media platform worldwide, with currently over 2 billion user accounts.

Interviews were conducted with representatives of six suppliers in India (SU-01 to SU-06): four chief executive officers, one operations manager, and one team leader. The six suppliers included two small domestic start-ups, three medium-sized domestic firms, and one medium-sized subsidiary of an MNE. Five of the firms are based in India, and one is located in the United States.

Our analysis of the content moderation labor process combines information from the management interviews with information from interviews with content moderators. In total, we interviewed nine content moderators (CM-01 to CM-09), all having permanent employment status for this study. CMs 05 to 08 were employed by a supplier firm but worked directly at the Indian subsidiary of an American video-sharing social media platform (GSM-02). In addition, we interviewed three content operators, two of whom worked at a domestic social media firm and one at a Chinese social media firm (CO-01 to CO-03). *Content operator* is a designation specific to domestic and regional social media firms; the job profile of a content operator extends beyond content moderation to other content-related tasks.

We faced several difficulties in accessing the target participants. We approached all of the workers on a popular professional networking website. We sent hundreds of workers a connection request, but only a few responded. Of these, a smaller number agreed to be interviewed. We applied a snowball sampling technique to contact further interviewees.

We conducted supplementary interviews with representatives of seven trade unions engaged in organizing IT workers. In order to understand how social media firms establish their content moderation policies, we also interviewed representatives from eight civil society organizations focusing on freedom of speech and online governance mechanisms in India.

### **The Position of Indian Firms in Content Moderation Value Chains**

Content moderation value chains are controlled by social media firms, mainly located in the Global North. These firms argue that they select their suppliers carefully, but they also emphasize the need for flexibility and the ability to quickly expand (but also contract) the volume of their outsourced content moderation business. The policy communications manager of a global social media company described the company's approach to choosing its outsourcing locations and partners as follows: "We work with a global network of partners, so we can quickly adjust the focus of our workforce as needed. For example, it gives us the ability to make sure we have the right language



expertise—and can quickly hire in different time zones—as new needs arise or when a situation around the world warrants it. These partners are carefully selected and reputable” (GSM-01).

Interviews with Indian suppliers show, however, the major role of price competition in the relations between social media firms and suppliers. The service level agreements of social media firms with their Indian suppliers are mostly price-based and project-oriented. Further, the two parties agree upon an overhead count (number of workers) before the onset of the project, and workers are employed by the contracted supply firms (SU-02, SU-04, SU-05). In part, Indian first-tier suppliers outsource certain projects to Indian second-tier suppliers, according to the managers of two small Indian companies (SU-05, SU-06).<sup>9</sup> The extent of such domestic outsourcing is unknown, however, as other managers of medium-sized content moderation firms stated that this would not be cost-effective (SU-02, SU-04, SU-05). Nevertheless, the manager of one supplier firm stated that freelance content moderators for its moderation projects are also sourced from external databases of Indian recruitment companies (SU-02).

There is a clear division of labor between social media firms and their content moderation suppliers. All product-oriented aspects such as training, moderation policies, and moderation software systems are managed by the social media firms. The human resources-related aspects such as wages, leave of absence, employment benefits, and other administrative tasks are managed by the supplier firms. Regarding recruitment and performance control, social media firms oscillate between the roles of supervision and direct intervention.<sup>10</sup>

Using the governance typology developed by Gereffi, Humphrey, and Sturgeon (2005), we can characterize the relationships between content moderation suppliers and social media firms in some cases as market-based and in other cases as captive. The social media firms exercise a high degree of power by dictating stringent standards in terms of technology and tools to be used for moderating content.

All content moderation suppliers interviewed for this study were very clear that there was no way of expanding their services or moving up the value chain. Certain possibilities for strengthening the position of Indian content moderation suppliers arise from the growing competition between the social media firms and from their interest in the Indian consumer market. The Indian social media landscape also benefits from growth of content in local languages. These developments have benefited both domestic and regional social media platforms, which primarily cater to consumers generating content in Indian vernacular languages.

The broadening of client firms through the inclusion of Chinese and domestic social media firms also presents opportunities to Indian content moderation firms, as it could

create a potential overlap of “multiple production networks” (Horner and Nadvi 2018) and reduce dependence on customers from North America and Europe.<sup>11</sup> India remains an important location for the Chinese technology company ByteDance, which owns both the video-sharing platform TikTok and the social networking platform Helo. However, a recent decision by the Union Government in India to ban 59 Chinese applications, including TikTok, in order to protect the country’s “national security and sovereignty” could limit these opportunities.<sup>12</sup>

### **The Content Moderation Labor Process**

The high power asymmetries between the lead firms and suppliers in content moderation GVCs strongly influence the labor process. In this section, we analyze the recruitment processes, work organization, and working conditions in this labor process. We also discuss the content moderators’ mobilities within and across moderation value chains, with a focus on their individual strategies for better working conditions and wages.

#### **Recruitment Process**

From the outset, it is useful to iterate that content moderation is not a standard business terminology similar to other job designations for consumer services in the IT sector. Instead, a careful analysis of advertisements for this work shows a diverse range of job titles, such as “system analyst,” “website administrator,” “process associate,” and “process executive.” The use of such generic terms can be attributed to several factors, including the diversity of firms offering content moderation services, the required skills, and the differing demands of global clients. These “multitudinous” job titles, as Roberts (2019) puts it, further obscure the landscape of content moderation from public visibility.

Mostly freshly graduated, the moderators we interviewed had all been encouraged by friends to apply for moderation jobs, often without knowing exactly what the work would entail. Those applying for work that included projects for global social media firms stated that it did not matter if the “job entailed BPO-styled working conditions” as long as the brands were well known in the employment market (CM-03, CM-06, CM-07, CM-08). However, agreements between the social media firms and suppliers require moderators to sign nondisclosure agreements, often even during the recruitment process. Along with working in opacity, moderators offer their labor in exchange for a mostly stagnant salary and few benefits under pressure of rising unemployment in the country.

The recruitment process is generally lengthy, with several rounds of interviews and assessments that aim to check the workers’ cognitive capacities and grasp of the required language for moderating content. In some cases (GSM-02), the social media firms directly

intervene in the process by conducting the final interview. In other cases, the suppliers undertake the complete recruitment process, with no participation by social media firms.

An important factor in the acceptance of relatively low wages and difficult working conditions is the composition of the content moderation workforce. The workers we interviewed for this study were young, below the age of 35 years, mostly male (75 percent of the sample), and unmarried. They held different educational qualifications such as engineering and technology, computer applications, management studies, media and communications, and education sciences. While we lack empirical data on where they undertook their higher studies, the workers came from both urban and rural regions of India. In many cases, content moderation was their first job, which they considered an entry point into the IT sector and a means of acquiring work experience.

### **Work Organization**

Moderation can start before user content is published on social media platforms (proactive filtering, which takes place in real time), or it may take place after it is uploaded (reactive filtering). Reactive filtering is often applicable in high-volume platforms such as those owned by social media firms GSM-01 and GSM-02 and depends on complaints made by external parties or users, who can flag or prompt review of content on the platform. In reactive filtering, two processes are involved: automated and manual moderation. Automated moderation entails automatic detection of user content matching the unique codes or hashes or digital fingerprints, resulting in deletion or approval of the content in compliance with the platform's policies. Child sexual abuse, revenge porn, and so forth are examples of content that are typically moderated via automated filters (e.g., Microsoft PhotoDNA). Usually, such content does not go into the manual queues.

However, a large amount of other content ends up in the queues of the moderators. Across all our cases, every moderator is assigned to a particular content queue, such as hate speech, copyright, or spam. The specific tasks of the content moderators depend on the policies of their customers. In most cases, the content moderators review massive amounts of user-generated content and make decisions to allow "flagged content" (i.e., content marked by users as offensive or unacceptable) on the platform, delete the content, or even ban the user. These decisions must be delivered at high speed and require intimate familiarity with the respective platform's policy guidelines. In some of the companies, there is a second layer of moderation, the quality analysis team, which checks the decisions of content moderators. In other cases, senior moderators may perform quality control.

In certain cases (CM-03, CM-04, CM-09), the content moderators were only tagging the content, since the policies of the social media platform they were working for did

not require the active deletion of content. Instead, they were required to tag the problematic content so that it became invisible to either the user or the country where this type of content was not allowed.

Through our analysis, we can observe that the content moderation process displays a strict work hierarchy, with moderators assigned to different levels based on their performance. The moderators are generally evaluated monthly against how many pieces of content they moderated and how many pieces they routed to colleagues in the other levels when they found it too difficult to decide themselves. In one of our cases, the moderator target (i.e., amount of content to be moderated) ranges from 2,500 to 5,000 pieces of content every month (CM-05, CM-06, CM-07, CM-08). The targets might differ in other cases, depending on the kind of content (video targets tend to be smaller), the size of the moderation workforce, and the demands placed by the social media firms.

We encountered diverging points of view regarding the automation of content moderation. On the one hand, several suppliers we interviewed were investing in automation of the labor process. Our research shows that they implement basic filters or more advanced automated technology for content moderation, depending on the requirements of the social media firms (SU-01, SU-04, SU-06). On the other hand, moderators working in two supplier firms that had already implemented the content moderation software provided by social media firms were vocal about the errors made by these tools (CM-03, CM-09). Rather than fearing replacement by automated moderation systems, they seemed more concerned about the low accuracy and extra work these tools create for them, such as correcting the automated suspension of genuine profiles of social media users (CM-03). A moderator working at a direct subsidiary of another social media firm (GSM-02) noted that while they have not encountered such problems, they have to ensure that such automation-generated errors are quickly resolved so as not to lose user trust in the social media platform (CM-08).

### **Working Conditions**

The most distressing element of content moderation work is the nature of the content that falls into the queues of the workers. Almost all workers recounted experiences of watching content involving pornography, assault, animal abuse, and live suicide. As we have described above, content moderators work in different content queues. Depending on the content queue (such as hate speech, nudity, self-harm, etc.) the frequency of the psychologically distressing content may be higher. In some content queues, such as news articles, the prevalence of distressing content is rare. Independent of the length of the training that they were given at the onset of their work, all moderators admitted that watching such disturbing content distressed them.<sup>13</sup> However, they believed that

one had to learn to adapt to such things. Bad management practices, long working hours, and lack of growth opportunities were cited instead as the main sources of their dissatisfaction with content moderation work. One moderator (CM-01), employed at the supplier firm SU-01, noted that he was not prepared to watch such content. SU-01 provided only one to two days of training to its employees at the outset of any content moderation project. An excerpt from the interview with CM-01 highlights their working conditions:

Sometimes I worked for 16 hours a day. After completing my shift, I used to go back home with an alcohol bottle and sleep. I tried also going with my friends to movies. I managed somehow. I had to look for better opportunities. I didn't have a laptop so I would borrow one from my friends, browse, try to learn something. From Coursera, I started learning neural networking. I had to ask Coursera to let me attend the course for free. I told them that I work for a small company. They accepted my request. It kept me busy and distracted.

Content moderators are cognizant of the multidimensional psychological impact of moderation on them, which extends beyond watching the distressing content. Their work is strictly timed and monitored through targets and "time punching." Failure to complete their tasks leads to disciplinary measures such as being issued statutory warnings, then being shifted to relatively easier (in terms of content complexity) work levels or even to another project, and eventually being "put on the bench," as content moderators described serving the notice period before their termination (CM-01, CM-06, CM-07, CM-08).

While conflict with management could range over several issues, such as work shifts, salary, or working hours, it also arose from lack of growth opportunities for moderators. Low wages, lack of skill development, and lack of promotion opportunities are the main reasons for the high attrition rate in this business. The suppliers note that high attrition rates are definitive of the IT sector in India and allow content moderators to participate in an expanding content moderation service market in the country (SU-01, SU-03, SU-04, SU-05). They affirm that constant recruitment drives nevertheless keep the turnover rate of workers high.

### **Climbing the Ladder or Creating Their Own Staircase:**

#### **Workers' Strategies for Change**

Content moderators are, however, not passive and defenseless. Roberts's (2019) argument that content moderators use their "high level cognitive functions" and "cultural competencies" is right. Moderators exercise agency using their cognition and cultural knowledge, albeit within a standardized and controlled labor process. Moderators are vocal about why they think global social media firms outsource moderation work to

India. Apart from the cost-effectiveness, they note that these firms need “localization experts” (C0-03, CM-06, CM-08). An anecdote from CM-06 who quit moderating for a global social media firm (GSM-02) after three years can be understood within this context: “I can tell you that these companies also outsource because they need the localization experts, because they are trying to capture the Indian social media market. If you look at this Chinese social media firm, they are hiring in India every day. You see, if you are good with the content, you know the local language, and you have the X factor in operations, you will get the job.” While a considerable level of cultural knowledge and understanding is necessary in all content moderation work, two factors are associated with an even higher demand for this knowledge.

The first factor is the type of projects and markets the Indian suppliers are serving. As we have already mentioned, the number of social media platforms in India has greatly increased. Expansion of regional and domestic social media platforms in the Indian market has created jobs for experienced moderators who are employed as content operators. Catering primarily to the Indian users who generate content in non-English and vernacular languages, the work of content operators constitutes a range of operations-related tasks such as user acquisition, user engagement, and developing content moderation policies, in addition to checking the quality of work by external moderators.

Content operators are employed directly at the regional and domestic offices of the lead firms’ offshoring units in India. “At least we are paid better here,” commented CM-08, who resigned from working as a content moderator in an Indian supplier firm and found new opportunities as a content operator working directly for a social media company. Better wages, however, do not always mean better work quality, as the content operators interviewed for our study decried the lack of training and growth opportunities (CM-06, CM-08).

The second important factor is the level of involvement of moderators in improving the social media firms’ content moderation policies. These policies are guarded by social media firms as “trade secrets” and function to “protect the client brand” (Roberts 2019). Soon after their recruitment, moderators at small- and medium-sized suppliers are trained for a few days, whereas large supplier firms organize training periods ranging from one to three months. The training provided by firms mostly focuses on formal content moderation policies. This policy training is mandatory in every service agreement between social media firms and suppliers and varies according to the needs of the clients and the capacities of the suppliers.

Content moderators note, however, that the rules and policies taught during the training are not all encompassing. Lack of clear policies on “edgy content” (CM-07), “newsworthy content” (CM-08), or even foreign language content (CM-05) make the

work difficult. Added to this are ambiguous policies regarding politicians and celebrities. In cases where there are no clear policies, moderators must create their own workflows, made up of a sequential set of steps in the work process. Giving an example of judging the intent of a user who posted a hacking video, CM-08, employed at a supplier working for a large social media firm (GSM-02), provides a glimpse into his new workflow:

This hacking video which once came in my queue was complicated. There were no clear rules on it. Therefore, I had to judge the intent of the user. The first thing I did was to check if it is a hacking tutorial or not. If it is a tutorial, then the policy requires me to allow it, but if the video is promoting hacking then I must delete it. Now how do you assess all this? I check the video title, the description, the video tags used, the user's channel, other videos posted by this user, and the nature of those videos. This is the way to make a new workflow. Takes a long time sometimes. No policy tells us how to check the intent of the user on the video.

Content moderators working directly for an Indian subsidiary of a social media firm (GSM-02) reported that they can influence the policy documents through what they call "ideation"—that is, providing ideas to make the policies better (CM-05, CM-08). This requires the moderators to record new solutions they create in the policy documents. Afterward, they make a presentation to their managers and demonstrate how such a policy change can be beneficial to the social media firm and its consumers. By playing a role in improving the policy documents, which are updated every few months, the content moderators can develop and demonstrate their expertise. We found this close involvement of content moderators in improving the moderation policies in only one case. In the other cases, the large distance between the social media firm and the supplier, as well as the high standardization of work processes, excluded this type of job enrichment.

In the absence of internal interest representation and career opportunities, moderators make use of the growing reliance of social media firms on the Indian labor market and construct their own "career staircases" (James and Vira 2012) across the different global, regional, and domestic firms engaged in the content moderation sector in India. At the same time, the Indian IT sector offers us concrete examples of collective struggles, especially concerning unpaid wages and layoffs. Interviews with trade unions show that following the 2008 global financial crisis, the effects on IT workers' job security started becoming noticeable. Both the Union for ITeS Professional (UNITES) and the Forum for IT Employees (FITE) were formed primarily in response to increasing layoffs at Indian IT firms (Noronha and D'Cruz 2020). However, there remains a lack of research about the struggles to unionize workers engaged in the content moderation value chains.

## Conclusion

The starting point of our analysis was the outsourcing of content moderation undertaken by global social media firms. These practices have largely remained hidden and continue to veil the relationships with suppliers and the working conditions of offshored content moderators.

The outsourcing of content moderation has created GVCs and global economic geographies characterized by strong power asymmetries. Located mainly in North America and Europe, the social media firms (i.e., the lead firms in the value chain) control the technological infrastructure as well as the rules and standards of the content moderation process, leaving very little room for upgrading on the side of Indian suppliers. Some cases can be described as captive value chain configurations, with the social media firms exerting tight control over all internal processes in the supplier firms. In other cases in our study, suppliers enjoyed a little more autonomy, and the value chain configurations could be described as market based. Even in these cases, the high standardization and codification of tasks as well as price competition limited the scope for supplier upgrading.

Understanding the important role of content moderation in the business model of social media firms enables us to situate the labor of content moderators, which is often contracted and located in internationally outsourced and offshored regions of the world. We have shown how opacity influences their labor process: applicants mostly apply for work having no information about the tasks they will have to perform, and they are strictly discouraged from talking about their work. At the same time, they work with very stressful content, in a strictly standardized workflow, and under enormous time pressure. It is not surprising that this labor model is characterized by a very high attrition rate and repeated reporting of health (psychological) damage.

Our analysis also shows that the knowledge and experience of workers remains important, despite the high standardization and technological control of the labor process. This was clear in one of our cases, where moderators had the possibility to influence the moderation policies developed by the social media firm. In most cases, however, content moderators have limited opportunities for skill development and growth. Moreover, integration within these GVCs does not lead to enhanced social protection and labor rights. Facilitated neither by the suppliers nor by the social media firms, content moderators must create their own paths for mobility and growth. Given the growing importance of social media markets in the Global South, Chinese and Indian social media firms are expanding and creating new employment opportunities for moderators. North American and European firms also increasingly need content moderators as localization experts. This is a source of labor power that can be used at



the individual level as well as a basis for organizing professional groups. Finally, firms and regulators have a responsibility to improve the conditions of this work, which is so critical for the functioning of social media platforms and for sustaining public discussion.

## Notes

1. See, for example, <http://gawker.com/5885714/inside-facebooks-outsourced-anti-porn-and-gore-brigade-where-camel-toes-are-more-offensive-than-crushed-heads>; <https://www.telegraphindia.com/7-days/guardians-of-the-internet/cid/1669422>; and <https://tech.economictimes.indiatimes.com/news/internet/meet-the-indian-warriors-who-watch-hours-of-beheadings-murders-gory-content-to-clean-the-internet/58901110?redirect=1>.
2. See <http://www.theguardian.com/technology/2017/dec/05/youtube-offensive-videos-journalists-moderators>.
3. See <https://www.forbes.com/sites/kalevleetaru/2018/09/08/is-social-media-content-moderation-an-impossible-task/>.
4. See <https://www.telegraphindia.com/7-days/guardians-of-the-internet/cid/1669422>.
5. The IT BPO sector in India includes a broad range of services supplied to different industries, such as health, finance, law, and technology.
6. These characteristics of market-based chains have also been observed by the GVC scholars in “modular” value chains, although the two differ regarding the complexity of transactions. We do not engage in a discussion on modular chains in the chapter because it is not relevant to our cases.
7. See, for example, <https://www.nytimes.com/2010/07/19/technology/19screen.html>; <https://www.wired.com/2014/10/content-moderation/>; <https://www.theatlantic.com/technology/archive/2019/06/facebook-and-youtubes-platform-excuse-dying/591466/>; <https://sz-magazin.sueddeutsche.de/internet/three-months-in-hell-84381>; <https://economictimes.indiatimes.com/tech/internet/indias-graduates-line-up-to-rid-facebook-of-inappropriate-content/articleshow/65676967.cms?from=mdr>; and <https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-inter-views-trauma-working-conditions-arizona>.
8. In the interest of protecting the anonymity of our target participants, we identify them by codes. DSM refers to domestic social media firm, SU refers to supplier, CM refers to content moderator, CO refers to content operator, and GSM refers to global social media firm.
9. First-tier suppliers provide their services directly to social media firms, and the second tier supplies these services to the first tier.
10. For some social media firms, direct intervention includes taking part in the recruitment process.
11. Interviews with content operators employed at domestic and Chinese social media firms in India show that these firms outsource moderation work to external suppliers in India. Having no direct communication with the external content moderators, our interviewees had no knowledge about who the moderators are and where they work. Therefore, the future potential of this outsourcing can only be estimated.

12. See <https://www.thehindu.com/news/national/govt-bans-59-apps-including-tiktok-wechat/article31947445.ece>.

13. The training provided by the companies focuses on formal content moderation policies and instructions on using the content moderation software. These trainings aim at instructing moderators on how to review content and make decisions regarding it. At the end of the training, the moderators must take a test, and, based on their results, they either are sent to different difficulty levels or retake the training.

## References

Ahmad, Sabrina. 2019. “‘It’s Just the Job’: Investigating the Influence of Culture in India’s Commercial Content Moderation Industry.” Master’s thesis, University of Oxford. Preprint, submitted in July. <https://doi.org/10.31235/osf.io/hjcv2>.

Batt, Rosemary, Virginia Doellgast, Hyunji Kwon, Mudit Nopany, Priti Nopany, and Anil da Costa. 2005. “The Indian Call Centre Industry: National Benchmarking Report Strategy, HR Practices, & Performance.” CAHRS Working Paper 05–07. Ithaca, NY: Cornell University, School of Industrial and Labor Relations, Center for Advanced Human Resource Studies.

Cairncross, Frances. 2001. *The Death of Distance: How the Communications Revolution Will Change Our Lives*. London: Orion.

D’Cruz, Premilla, and Ernesto Noronha. 2008. “Doing Emotional Labour: The Experiences of Indian Call Centre Agents.” *Global Business Review* 9 (1): 131–147.

D’Cruz, Premilla, and Ernesto Noronha. 2013. “Hope to Despair: The Experience of Organizing Indian Call Centre Employees.” *Indian Journal of Industrial Relations* 48 (3): 471–486.

Gereffi, Gary, John Humphrey, and Timothy Sturgeon. 2005. “The Governance of Global Value Chains.” *Review of International Political Economy* 12 (1): 78–104.

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

Gillespie, Tarleton. 2018b. “Regulation of and by Platforms.” In *The SAGE Handbook of Social Media*, edited by Jean Burgess, Alice Marwick, and Thomas Poell, 254–278. London: SAGE.

Graham, Mark, and Mohammad A. Anwar. 2019. “The Global Gig Economy: Towards a Planetary Labour Market.” *First Monday* 24 (4), April 1. <https://doi.org/10.5210/fm.v24i4.9913>.

Grimmelmann, James. 2015. “The Virtues of Moderation.” *Yale Journal of Law and Technology* 17 (1): 42–109. <https://digitalcommons.law.yale.edu/yjolt/vol17/iss1/2/>.

Horner, Rory, and Khalid Nadvi. 2018. “Global Value Chains and the Rise of the Global South: Unpacking Twenty-First Century Polycentric Trade.” *Global Networks* 18 (2): 207–237. <https://doi.org/10.1111/glob.12180>.

James, Al, and Bhaskar Vira. 2012. “Labour Geographies of India’s New Service Economy.” *Journal of Economic Geography* 12 (4): 841–875.

Klonick, Kate. 2017. "The New Governors: The People, Rules, and Processes Governing Online Speech." *Harvard Law Review* 131 (6): 1599–1670. [https://harvardlawreview.org/wp-content/uploads/2018/04/1598-1670\\_Online.pdf](https://harvardlawreview.org/wp-content/uploads/2018/04/1598-1670_Online.pdf).

Lakhani, Tashlin, Sarosh Kuruvilla, and Ariel Avgar. 2013. "From the Firm to the Network: Global Value Chains and Employment Relations Theory." *British Journal of Industrial Relations* 51 (3): 440–472.

Medeiros, Ben. 2017. "Platform (Non-)Intervention and the 'Marketplace' Paradigm for Speech Regulation." *Social Media + Society* 3 (1): 1–10. <https://doi.org/10.1177/2056305117691997>.

Nadeem, Shehzad. 2011. *Dead Ringers: How Outsourcing Is Changing the Way Indians Understand Themselves*. Princeton, NJ: Princeton University Press.

Noronha, Ernesto, and Premilla D'Cruz. 2006. "Organising Call Centre Agents: Emerging Issues." *Economic and Political Weekly* 41 (21): 2115–2121.

Noronha, Ernesto, and Premilla D'Cruz. 2020. "The Indian IT Industry: A Global Production Network Perspective." Working Paper no. 134. Berlin: Berlin Institute for International Political Economy, Hochschule für Wirtschaft und Recht Berlin. [https://www.ipe-berlin.org/fileadmin/institut-ipe/Dokumente/Working\\_Papers/ipe\\_working\\_paper\\_134.pdf](https://www.ipe-berlin.org/fileadmin/institut-ipe/Dokumente/Working_Papers/ipe_working_paper_134.pdf).

Peck, Jamie. 2017. *Offshore: Exploring the Worlds of Global Outsourcing*. Oxford: Oxford University Press.

Ponte, Stefano, and Timothy Sturgeon. 2014. "Explaining Governance in Global Value Chains: A Modular Theory-Building Effort." *Review of International Political Economy* 21 (1): 195–223.

Remesh, Babu. 2008. "Work Organisation, Control and 'Empowerment': Managing the Contradictions of Call Centre Work." In *In an Outpost of the Global Economy: Work and Workers in India's Information Technology Industry*, edited by Carol Upadhyia and A. R. Vasavi, 235–262. New Delhi: Routledge.

Roberts, Sarah T. 2016. "Digital Refuse: Canadian Garbage, Commercial Content Moderation and the Global Circulation of Social Media's Waste." *Wi: Journal of Mobile Media* 10 (1): 1–18.

Roberts, Sarah T. 2017. *Content Moderation*. In *Encyclopedia of Big Data*, edited by Laurie A. Schintler and Connie L. McNeely, C: 1–4. Cham, Switzerland: Springer International.

Roberts, Sarah T. 2019. *Behind the Screen: Content Moderation in the Shadows of Social Media*. New Haven, CT: Yale University Press.

Schrage, Stephanie, and Dirk Ulrich Gilbert. 2019. "Addressing Governance Gaps in Global Value Chains: Introducing a Systematic Typology." *Journal of Business Ethics* (January): 1–16.

Shepherd, Tamara, Alison Harvey, Tim Jordan, Sam Srauy, and Kate Miltner. 2015. "Histories of Hating." *Social Media + Society* 1 (2): 1–10. <https://doi.org/10.1177/2056305115603997>.

Smith, Chris, and Paul Thompson. 1998. "Re-Evaluating the Labour Process Debate." *Economic and Industrial Democracy* 19 (4): 551–577.

Taylor, Phil, and Peter Bain. 2005. "'India Calling to the Far Away Towns': The Call Centre Labour Process and Globalization." *Work, Employment and Society* 19 (2): 261–282.

Tenhunen, Sirpa. 2018. *A Village Goes Mobile: Telephony, Mediation, and Social Change in Rural India*. New York: Oxford University Press.

Thompson, Paul. 2003. "Fantasy Island: A Labour Process Critique of the 'Age of Surveillance.'" *Surveillance & Society* 1 (2): 138–151.

Upadhy, Carol. 2010. "Taking the High Road? Labour in the Indian Software Outsourcing Industry." In *Labour in Global Production Networks in India*, edited by Anne Posthuma and Dev Nathan, 300–320. New Delhi: Oxford University Press.