"It's Just the Job": Investigating the Influence of Culture in India's Commercial Content Moderation Industry

Thesis submitted in partial fulfilment of the requirements for the degree of MSc in Social Science of the Internet at the Oxford Internet Institute, University of Oxford

Candidate: Sabrina Ahmad
Submitted: July 2018 (Trinity Term, 2018)
College: Balliol
Oxford Internet Institute
University of Oxford
Word Count: 14, 459

Abstract

As user-generated content (UGC) on social media becomes near-ubiquitous, the need to review potentially insidious text, images, and video before and after they surface has resulted in the emergence of a new industry in global digital labour: commercial content moderation (CCM). A psychologically taxing job, a handful of scholars and journalists have studied the emotional burden for content moderators in the West, many of whom are predominantly based in Silicon Valley. However, much of this work is increasingly outsourced to the Global South, namely in India and the Philippines. Despite the fact that these overseas moderators filter UGC from Western countries, where the cultural norms on topics like hate speech and sexuality differ from those in Eastern countries like India, the academic research to date has neglected to study foreign content moderators and the ways in which their culture may impact their decision-making processes in reviewing UGC. This thesis takes a qualitative approach using semi-structured interviews to study content moderators in India with the objective of understanding not only the country's booming and rapidly-evolving content moderation industry, but more essentially the subjectivities of Indian moderators and how their culture might influence their work. In addition to presenting novel findings in the general field of CCM, this research ultimately demonstrates that many of the assumptions about Indian content moderators – primarily that they are unable to separate their own cultural upbringings from the US-centric policy guidelines they are meant to uphold – are erroneous and in need of revisiting.

Table of Contents

Abstract	
1. Introduction.....	1
2. Literature Review	4
2.1. Commercial Content Moderation: An Emergent Field.....	4
2.2. The Inefficacy of Automation in Content Moderation.....	6
2.3. Obfuscating the Moderator: Platform Incentives and the Political Economy of CCM.....	7
2.4. Outsourcing Moderation: Culture and Colonialism in Global Digital Labour	8
2.5. The Imperative for Studying Indian Content Moderators.....	11
3. Methodology	12
3.1. Overview	12
3.2. Recruitment	12
3.3. Interviews	14
3.4. Thematic Analysis.....	16
3.5. Ethics	17
3.6. Limitations	18
4. Findings.....	19
4.1. Diversity of Projects and Clientele.....	20
4.2. Nature of the Work: Rote Review Versus Analytical Thinking	20
5. Analysis	23
5.1. The Structural Dimension: The Gendered Division of Labour in India's Content Moderation Industry	24
5.2. The Technical Dimension: Cultural Competence Training for Indian Content Moderators	29
5.3. The Subjective Dimension: Unearthing Perceptions Amongst Indian Content Moderators	34
6. Conclusion.....	40
References.....	42
Appendix A: CUREC Approval	46
Appendix B: Interview Schedule.....	47
Appendix C: Qualitative Coding Schedule	48

1. Introduction

Besides the occasional disturbing image or video, the majority of us are insulated from material on the Internet that confronts us with the darkest depths of humanity. This is not, however, a result of the inherent positivity of the world, but instead a careful process of curation known as content moderation performed by platforms big and small. And with the sheer volume of content generated on social networking websites today, content moderation is becoming an increasingly critical process that has recently attracted the attention of Internet scholars. It has also garnered the attention of the general public who endeavour to better understand who intervenes in the decisions about what constitutes admissible material on the websites they visit each and every day. Algorithms have become adept at filtering what would be considered offensive content online, but they lack the sophistication to effectively distinguish nuances in homophobic, racist, sexist, or violent material inappropriate for public consumption (Feamster, 2018). As such, this taxing and psychologically burdensome work gets outsourced to humans, many of whom are poorly-paid content moderators operating out of business process outsourcing (BPO) centers throughout India and the Philippines. Given contemporary social media usage, content moderators play a fundamental though oft-ignored role in keeping platforms safe for billions of users.

One challenge arising in the context of such outsourcing is that moderators in the Global South are accused of being unfamiliar with the norms and conventions that govern these US-centric platforms. But how, and to what degree, do incongruent cultural norms between the West and the Global South influence the decisions that a content moderator makes? While these platforms have written guidelines or “community standards” to which overseas content moderators must refer, these standards are difficult to decipher and interpret. Much of what should be kept or discarded is ultimately left to individual judgement, and is thus allegedly ultimately implicated in some form of bias.

The fact that there are individuals with their own prejudices and beliefs deciding what content is moderated online reveals a fascinating area of research that warrants a deeper and more thorough investigation. Specifically, virtually no academic literature addresses the work of overseas content moderators in India; while a handful of scholars have examined the emotional labour for content moderators in the West, focalizing primarily on the emotional and physical exhaustion that such workers endure (Gillespie, 2018), none have engaged with the intersection of Indian culture vis-à-vis content moderation. This is a critical issue given that overseas moderators sift through user-generated content (UGC) from Western countries, where the cultural norms differ from those in India (Lewis, 2006). Indeed, all scholarly studies in the realm of content moderation have adopted a decidedly Western focus, while only in the most cursory fashion mentioning that a large proportion of content moderators work out of India.

This important gap in the literature has thus inspired the following research question: *In what ways do the cultural beliefs of Indian content moderators impact how content is moderated on online social networking sites?* This thesis takes a qualitative approach using semi-structured interviews to study content moderators in India with the objective of understanding not only the country's booming and rapidly-evolving content moderation industry, but more essentially the subjectivities of Indian moderators and how their culture might influence their work. For the purposes of this research, "culture" refers to an intertwined system of values, attitudes, beliefs, and norms that give meaning to both individual and collective identity (Adler in Shah, 2004). This research also demonstrates that all previous academic study on commercial content moderators treats the experience of content moderators in North America as somehow universally applicable to moderators working in other parts of the globe, which is unrepresentative and an injustice to the Indian moderators who work in dramatically different conditions and contexts from their US-based counterparts. This research also presents novel findings in the realm of commercial content moderation: where previous work has concentrated primarily on the "platform giants" such as Google, Facebook, and

YouTube (Roberts, 2016; Gillespie, 2018; Klonick, 2017), this thesis shows how moderation is performed by Indian moderators in a more analytical, rather than purely rote, manner for niche websites ranging from meme repositories to Belgian online dating websites. It also analyzes three dimensions of content moderation in India – how the work is experienced along structural, technical, and subjective lines. Structurally, because content moderation in India is divided based on the perceived capabilities of each gender; technically, in the cultural assimilation process Indian moderators undertake to service their global clients; and, subjectively, in how Indian moderators experience moderation entirely differently from their American colleagues.

Most importantly, this research rejects the multitude of assumptions levelled against Indian content moderators in the literature. Rather than culture influencing the decision-making processes of these moderators, culture structurally influences the distribution of labour in Indian commercial content moderation. Where the literature attributes mistakes from Indian content moderators to unsubstantiated claims of “cultural bias,” this research illustrates that the training processes of these workers are exhaustive and technical. Finally, where Indian moderators are seen as victims to a revolving door of rancorous content to moderate, this thesis reveals the subjective experiences of Indian moderators and shows that they are meticulous, hard-working, and take pleasure in their responsibility to keep the Internet safe for billions of individuals all around the world.

In terms of structure, this research paper will first begin with a brief overview of the literature on content moderation written to date. It will then proceed with an explanation of the qualitative methodology used, followed by a presentation of the findings of the research and an analysis. It will then conclude with a summary of the key contributions of the research as well as recommendations for future study on content moderation.

2. Literature Review

2.1. Commercial Content Moderation: An Emergent Field

Sarah T. Roberts was the first to coin the term “commercial content moderation” (CCM) and defines it as the “organized practice of screening user-generated content (UGC) 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” (Roberts, 2017, p. 1). This is a largely incomplete definition, as policy guidelines at platforms who moderate UGC are not predicated solely on static features like locality or jurisdiction, but also evolving economic incentives, user bases, and audience tastes. The definition also fails to acknowledge the complex ecosystem of content moderation, which is a dynamic interplay between automated tools, human moderators, policy experts, legal consultants, volunteer reviewers, and more (Gillespie, 2018). Roberts also first laid the foundation for the importance of content moderators through the angle of preserving the economic viability of platforms: platforms must protect their brand reputation and image to continue to make revenue, and as such cannot allow for inappropriate content to unduly harm or upset users. Klonick (2017) points out in her overview of moderation practices at Facebook, Google, and YouTube that moderation can take on two different forms: *ex ante* moderation, in which content is screened at the moment between a user uploading a piece of content, and its publication; and *ex post* moderation, in which the UGC is reviewed reactively after having been “flagged” through the platform’s reporting mechanism. Interestingly, Klonick claims that *ex ante* moderation is largely performed by machines through an automatic process involving algorithmic screening, absent of active human decision-making: as this thesis will later demonstrate, this is inaccurate, as the research participants who work as content moderators in India attested that all of the work they did on behalf of clients was what they termed “pre-moderation,” or what Klonick here has designated *ex ante* moderation.

It is worth noting that the bulk of study into the content moderation industry comes not from scholarly research in the field, but from investigative journalism. Adrian Chen (2014) was among the first to write about content moderators, speaking with both foreign-contracted moderators in the Philippines as well as domestic, Bay Area-based moderators for a piece in *Wired*. Chen then followed this with a documentary co-directed with Ciaran Cassidy in 2017 that focalized on the training processes of Indian content moderators at a Bangalore-area company called Foiwe. It was thus in 2017 that a litany of pieces was written on the underbelly of the content moderation industry, running the gamut from widely-read publications such as *The Wall Street Journal* to more technology-specific digital news microsites like *The Verge*. However, a common thread running through all of these publications is the voyeuristic fascination with the gruesome realities of content moderation: emphasized is the notion that moderators regularly remove images and videos of “bestiality,” “beheadings,” “child pornography,” and more (Chen, 2014; Solon, 2017; Madrigal, 2017). Moreover, these pieces have predominantly concentrated on the subjective experiences of Western content moderators working in contracted capacities for large platforms like Facebook, Microsoft, and YouTube, underscoring their meagre salaries, psychological trauma, and the paucity of mental health resources available to them.

The pressing need for a theoretical framework through which to understand and regulate content moderation has attracted inspective efforts from even the United Nations. The UN’s 2018 Special Rapporteur report, produced by David Kaye, sought to understand platform content regulation from a human rights perspective from two positions: the first being on the side of government, for example where State obligations to establish punitive frameworks for the rapid removal of UGC undermine freedom of expression. Secondly, the report considered issues from the company content moderation side, in terms of policy and guideline development. The ultimate aim of the report was to reconcile these two elements of content moderation in a way that “puts human rights at the very centre,” and by this, he means the

rights of *users* (Kaye, 2018, p. 3). The role of human content moderators is mentioned only in passing: Kaye notes that fostering accurate content moderation practices requires that companies strengthen protections for human moderators that are consistent “with human rights norms applicable to labour rights,” but goes no deeper than this (Kaye, 2018, p. 18). In sum, Kaye’s report has endeavoured to understand how human rights, more specifically freedom of expression, may be championed in content moderation.

2.2. The Inefficacy of Automation in Content Moderation

In an age where accelerations in the research and development of artificial intelligence (AI) tools are at an all-time high, one might question whether there is even a need to continue to employ human content moderators. In his congressional questioning in April 2018, Mark Zuckerberg claimed that within the next five to 10 years, AI would be able to solve many of the company’s most pressing problems, including the proliferation of hate speech and terrorist propaganda (Harwell, 2018). However, the empirical evidence shows that machine learning, even at the most cutting-edge companies, is far from possessing the level of prowess necessary to be able to provide content moderation services with the same degree of accuracy and refinement as human moderators.

The Center for Democracy and Technology (2017) makes the case for why human moderators will play a pivotal role in content moderation for the foreseeable future: machines still have limited ability to parse the nuanced meaning of human communication and to detect the intent and motivation of the speaker, which could lead to disproportionate censorship and the biased enforcement of laws. Furthermore, the meaning of language is extremely dependent on contextual factors like tone, speaker, and audience: a human review of flagged content is still required for nuance, an example being when Instagram’s DeepText hate speech filter accidentally identified “fag,” a cigarette in British vernacular, to be inflammatory. Feamster (2018) holds that the future of AI and online content moderation will always, to some degree,

involve humans: detection algorithms are only as good as their data, and when this data continues to be biased and mislabeled, AI will continue as merely an assistive technology to help triage, scale, and improve human effectiveness and efficiency when executing decisions in content moderation.

2.3. Obfuscating the Moderator: Platform Incentives and the Political Economy of CCM

The economic incentive for platforms to engage in content moderation practices with the use of human moderators is highly compelling. As Roberts (2018) argues, moderation activities are primarily undertaken with a view to protect and enhance a platform's advertising revenue and are matter of brand protection. Klonick (2017) also observes that platforms are economically responsive to the expectations of its users who drive attention, clicks, and ad revenue, and as such, are disincentivized on a financial level from allowing anti-normative content out of fear it will repel large segments of its userbase. Not only are crowdworkers a cost-effective and flexible labour force that can perform detection and classification in a way machine learning techniques cannot, they are also able to absorb the psychological toll of doing so and remain hidden and removed in such a way so as to be detached from the legal protections associated with traditional employment (Gillespie, 2018). Thus, it can be seen that human moderators are not only an efficient means of policing content online, but they remain largely concealed from public scrutiny due to the unsavoury nature of the work. Obfuscating moderators consequently renders platforms as seemingly objective in the public imagination, conjuring an illusion that moderation is done using scientific, algorithmic tools driven by mechanical behaviour (Roberts, 2018).

In terms of the political economy of content moderation, moderation can be seen as a critical process in the social media ecosystem precisely because it has the power to determine what social reform movements, collective action campaigns, and information gain exposure and visibility (West, 2017). Content moderators challenge the concept of the Internet as a

democratic and free speech zone by working at the junction of what content gets deleted or stays put, thereby influencing the pipeline and social virality of UGC. Moderators also work a taxing job and feel pressure beyond the directive to simply remove objectionable content. For example, content that is sensationalistic or distasteful may in many cases be content that drives the most clicks, thereby situating moderators in a paradoxical role where they must balance a company's profit motive with the demands for brand protection (Roberts, 2016). Furthermore, moderators are often not given context when screening UGC: users are expected to weigh competing values and demonstrate cultural sensitivity, but instead the meaning behind images and posts are compressed and lost when there is an endless queue of them to moderate within minutes (Gillespie, 2018). It is for these reasons that moderators are hidden behind the scenes – opacity hides not only the fact of selection, but the market-driven values motivating that selection and the chaotic, ad-hoc nature of the process in general (Roberts, 2018). Platforms are reticent to draw attention to moderators precisely because it would reveal how few of them actually do the work, the unpleasant labour practices that underlay them, and the lack of automation implicated in the process which suggests the job is largely subjective (Gillespie, 2018). When competing cultures between a platform's policies and the moderators who implement them are thrown into the mix, the process becomes all the more discomfiting.

2.4. Outsourcing Moderation: Culture and Colonialism in Global Digital Labour

A crucial characteristic in the global chain of CCM is that moderators must be experts in a website's presumed audience: this poses a challenge to moderators based abroad who are required to be linguistically and culturally competent in English, oftentimes their second language (Roberts, 2014). That is, moderators working in countries such as India are removed geographically and culturally from where the content is generated: they work in call center environments that cater to Western business needs and are expected to conform to the cultural norms that match their clientele, which may be at odds with their own (Roberts, 2014). Thus,

when CCM work is outsourced to other parts of the world, an additional challenge is created wherein foreign moderators must learn what constitutes racist or sexually explicit material in another culture (Roberts, 2016). Furthermore, boutique American and European content moderation and social media management firms attempt to leverage this alleged “mastery” of Western cultural values as their unique selling proposition when competing with content moderation firms in the Global South. For example, a company formerly known as Caleris used as their tagline, “Outsource to Iowa. Not India” (Roberts, 2014, p. 32). Caleris’ co-founder Sheldon Ohringer claimed that his Iowan employees were more culturally aligned with Americans’ views of what is appropriate when it comes to UGC. Indeed, many of the content moderation firms from the Global South will impress upon prospective clients their proficiency with English and their familiarity with Western issues. For example, one Indian content moderation company claims that its moderators are familiar with “all relevant and current legal issues, copyright concerns, Child Safety considerations and other practices across the globe” (Infoesearch, 2018). Another Indian content moderation company similarly boasts about its knowledgeable and cosmopolitan team, who are “multi-lingual and highly qualified specialists” (Foiwe, 2018). Nearly all of these content moderation companies in India also tout their 24/7/365 availability and ability to work around the clock in order to protect a client’s brand reputation online, wherever in the world they may be based and at whatever time.

The growing phenomenon of globalized outsourcing in India also has implications for the moderators who perform this work. It comes as very little surprise that contracting to countries in the Global South is so popular: the labour in India most importantly is comparatively far cheaper than its European or Western counterparts (Roberts, 2016). What’s more, technological hubs in sprawling Indian urban metropolises like Hyderabad and Bangalore already possess the needed IT infrastructure to support robust, 24/7 content moderation services. India has an established and celebrated history with BPO call centers, being the world’s second-largest English-speaking country with an educated and technically literate

workforce that alone processes 50% of the world's offshored business (Gopelwar, 2012). Finally, a content moderator working both geographically and bureaucratically through a series of contracts removed from the platform allows for plausible deniability on behalf of the company. If there are any repercussions from moderating, such as a moderator in India being disturbed by something they have seen, he or she is distanced and unable to bring forth any complaints in a meaningful way (Roberts, 2016). In what might be considered still more prohibitive, these workers are prevented from speaking about their everyday working conditions or grievances by being required to sign non-disclosure agreements (NDAs).

This outsourcing to the Global South is hotly contested in academic spheres and has most recently been studied through the perspective of the perpetuation of colonialist practices in global digital labour. Graham, Hogan, Straumann, and Medhat (2014) note that the Internet's uneven geographies in the international division of labour constitute a central element of the platform economy. Internet moderation in particular attracts ambitious labour: the low-wage, secret review of images, video, and text is a less creative and more distressing form of digital work that is associated oftentimes with "data janitors," and "click farmers" (Casilli, 2016, p. 18). The books and documentaries on CCM follow this rhetoric, as is evidenced by titles such as Gillespie (2018)'s *Custodians of the Internet* and *The Cleaners*, a newly-released documentary about the content moderation industry in the Philippines. Nakamura (2014) adds that in the present context of global connectivity, such labourers are spatially hidden and deliberately consigned to out-of-the-way places where capitalist dynamics actively conceal the ways in which race and gender are key elements of digital platform production. The same does not necessarily apply to US-based workers, who are given affordances their Indian counterparts are not, like daytime work hours, regulatory frameworks, and job security (Casilli, 2016). The whole outsourcing industry has even been referred to as a "digital sweatshop" (Graham, Wood, Hjorth, & Lehdonvirta, 2016), where a new class of workers from developing countries are caught betwixt representing a workforce necessary for the global digital economy, but who

simultaneously get relegated to the margins without welfare benefits or a clear career trajectory. Thus, such economic strategies and the obvious efforts to overcome regulation in developing countries to serve the interests of Western platforms have overtones of colonialism, though many argue as to whether these are veritable *manifestations* of colonialism (Casilli, 2016).

2.5. The Imperative for Studying Indian Content Moderators

The literature, as has been illustrated thus far, concentrates overwhelmingly on the experiences and context of content moderation in the West. The little research that does acknowledge the challenges of content moderators abroad only does so sparingly, and with that, from an ethnocentric perspective. The following citations come from Roberts (2014)'s interviews with moderators based in the United States who worked with contracted colleagues in India:

Max*: The culture is so dramatically different when you are trying to apply a policy that is based on a United States or Western Europe culture [pause] it doesn't work. So they tried outsourcing to India before I was there, and it was such a disaster that they just canceled it and said we have to do this all in-house . . . [the Indian contractors] would be removing pictures of people at the beach with their family in bikinis, because they are like, that's not appropriate for public consumption.

Josh: If it's going to be hate speech or if it's conspiracy talk, so you actually have to watch those. Where as if it is pornography or spam it is instantly recognizable. So we took all of those videos, which are most of our videos, I would say 60-70% of our stuff is spam and porn, and that all went to our team in India.

Roberts (2014) admitted that the CCM workers from Silicon Valley whom she interviewed “cast a downward eye” towards globally dispersed CCM workers from India and the Philippines due to their inability to understand American nuance or language conventions (p. 164). It is through this stereotyping, imperialist gaze that much of the research on content moderation – by white, university-educated North American academics about predominantly

* No real names have been used in this paper; all quotes are attributed to a pseudonym.

white, university-educated North American content moderators – is constructed. Indeed, in Jessie Daniels (2012)’s critique of race and racism in Internet studies, it has been clear that not only is race built into the Internet industry, but that more often than not, the onus and burden of studying racialized groups on the Internet has been left to ‘minority participants’ – that is, researchers who are people of colour (p. 707).

This is still very much the case today and among the primary motivations for conducting this research project. As a person of colour, it was striking to note that in all the discussions about content moderation, little attention or concern was paid to the moderators who do this work abroad. What little attention that has been given to content moderators in the Global South takes on an air of superiority. Moreover, this inaccuracy and inefficiency is blamed on the “culture” of these foreign moderators, which, as seen in the bikini example, is implied to be regressive and archaic. Thus, this research seeks to investigate in a more thorough and methodologically rigorous manner whether claims of cultural influence adversely impacting the ability of Indian moderators to perform content moderation accurately holds credibility.

3. Methodology

3.1. Overview

This research employs a qualitative approach with a purposive snowball sampling method and semi-structured interviews. As such, this methodology section will first describe the recruitment process for contacting candidates; next, it will outline the rationale for the use of interviews as a method of data collection best suited to answer the research question; and, finally, it will describe the ethical considerations to this approach as well as its limitations.

3.2. Recruitment

The sampling for this study used purposive and snowball methods to recruit three specific groups of individuals: content moderators working in India, C-Suite executives at Indian

content moderation companies, and policy managers based in Silicon Valley who craft the moderation guidelines to which moderators abroad adhere. The first group, Indian content moderators, were chosen because they are the immediate subjects of interest for this research. The C-Suite executives at the Indian content moderation firms were interviewed to complement the perspectives of the Indian moderators and to contextualize the environment in which the moderators work; what are the aims and corporate philosophies of the company, what are their training procedures, and how do these features influence the work of the moderators themselves? Finally, the US-based policy managers were interviewed to give a holistic portrait of the content moderation industry, as well as to better understand the allegations raised against Indian moderators and how contracting to India in the global content moderation chain may impact accuracy and quality due to contrasting cultural norms. All three groups also serve to answer the research question, by examining from different angles how culture might influence the moderation processes of Indian content moderators.

As Roberts (2014) indicated in her dissertation, the most substantive barrier to conducting a study on content moderators is to locate participants. Roberts wrote that she anticipated that “access to workers” would be an “ongoing challenge” (Roberts, 2014, p. 59). As has already been established, content moderators are submerged beneath a quagmire of NDAs and an air of secrecy; as such, they are not out in the open, and the individuals who *are* open about their positions in the industry are reluctant to discuss their work due to potential legal repercussions. As the content moderators of interest to this particular study were based in India, the difficulties in finding them were exacerbated by geographic distance, differences in time zones, and lack of an existing network with mutual connections from which to draw.

In order to ensure that participants were found within the allocated timeframe, a multi-pronged approach was executed. Firstly, a list of all the major content moderation companies in India was compiled, and the general inboxes at each of these companies systematically emailed with information about the research and a request to speak with a moderator. Parallel

to these recruitment efforts, private messages on LinkedIn were sent to individuals whose titles were “content moderator” and “systems analyst” at major Indian content moderation companies. Other efforts included reaching out to journalists who have produced documentaries on content moderation in India. Finally, requests for connections were made with peers at the Oxford Internet Institute (OII), specifically those who had worked in content removals.

The method that yielded the most fruitful results came from privately messaging Indian content moderators on LinkedIn. The explicit use of Oxford in the subject heading, it should be noted, was a highly intentional one: as a result of colonialism, the British Raj left both a social and cultural influence on many Indians, with commonalities between the British including a reverence for “cricket, tea... [and the] Oxford and Cambridge elite” (Lewis, 2006, p. 435). In sum, the name recognition of Oxford and its cultural status among middle-class Indians likely played a significant factor in the response rate.

Ultimately, the second- and third-most effective methods for soliciting research participants came from personal connections and journalists. Once two to three initial interviewees were successfully contacted through LinkedIn messaging, and C-Suite executives as well as Silicon Valley-based policy managers were found with the help of journalists and the OII network, snowball sampling facilitated the acquisition of the remaining interview sample.

3.3. Interviews

Ten semi-structured interviews over Skype were conducted in total; six with Indian content moderators, two with C-Suite executives at Indian content moderation firms, and two with former Silicon Valley Trust and Safety policy managers with previous experience in crafting moderation guidelines for large platforms, and who managed teams in India administering

those guidelines. The interviews were 30 to 45 minutes long, audio-recorded, and partially transcribed (see Appendix B for interview schedule).

In terms of justification for this methodology, interviews are a dyadic form of interaction used in qualitative research and are useful for exploring a general topic area in order to glean new insights (Gibson & Hua, 2016). Since the research question is concerned with the ways in which the cultural beliefs of Indian content moderators may impact decision-making when it comes to moderating online UGC, justifications and explanations as to why a moderator has acted in a particular manner are needed. This is best served through a qualitative, interview-based approach since more probative questions can be asked that solicit in-depth answers conducive to inductive methods of analysis such as thematic content analysis (Braun, Clarke, & Rance, 2014). Thus, interviews were used for this research to gain trust and proximity to moderators in order to yield valuable information about their experiences, “subjective understandings,” (Seidman, 2006, p. 10) and how their cultural beliefs may impact their work. The interview is also a method through which one learns about others’ stories (Seidman, 2006) – content moderators, whose contributions to keeping our online world safe often go unacknowledged, have important stories to tell about the subversive ways in which they do so. The choice of semi-structured interviews as the method of qualitative inquiry for this research was also used for several reasons. As has been previously discussed, the nature of the research question is highly sensitive: semi-structured interviews allow for fluidity where it is possible to pursue different avenues depending on what each participant is comfortable discussing or what he or she finds interesting. Additionally, the type of iterative questioning in semi-structured interviewing, where the interviewer returns to or rephrases earlier questions, augments credibility by making it possible to detect contradictions or any confusion and resolve such issues (Shenton, 2004).

The number of interviews conducted was chosen for specific reasons. While the number of interviews considered sufficient varies considerably amongst qualitative research practitioners,

what is agreed upon is that the number should align with the researcher's objectives and resources (Baker & Edwards, 2012). The objective of this research was to understand the cultural impact of overseas content moderators in India reviewing Western UGC, and this research was conducted within the span of three months from the UK – it is, in this author's own estimation, a timely and insightful analysis of the globalized content moderation industry and it achieves its intended goal despite geographic and logistical barriers. As Adler and Adler remark, when subjects are “easy to find and plentiful,” researchers may gather more interviews from them (in Baker & Edwards, 2012, p. 9). Indian content moderators by the very nature of their work are hidden from public sight, an explicit decision made by the platforms who employ them. Thus, the number of interviews obtained remains impressive despite the obstacles delineated. Finally, more interviews become redundant when no new themes can be identified (Brannen in Baker & Edwards, 2012). Since theoretical saturation was obtained after ten interviews, it was unnecessary to seek more.

3.4. Thematic Analysis

The interview method in this research was used precisely because it produces rich qualitative data ripe for thematic analysis. A “theme” in qualitative research refers to a phrase or sentence that identifies what a unit of data is about, or what it means (Saldaña, 2009). In terms of a systematic manner of approaching thematic analysis, it was first important to recognize cues that demarked an emerging theme, such as repetition (Ryan & Bernard, 2003). Indeed, themes became crystallized as certain phrases or topics appeared recurrently across transcripts. Another cue that indicates the possibility of an interesting theme is the use of metaphor (Ryan & Bernard, 2003): this contributed to the “structural dimension” theme, or the gendered division of labour, which was partly inspired by one participant discussing men and women working together as a “battalion.” Furthermore, the themes identified should be relevant to the researcher's objectives – in this instance, the “subjective” theme is pertinent

because, in addition to being repeated throughout transcripts, the research question endeavours to understand how culture impacts the decision-making processes of Indian content moderators which is in itself subjective. The “structural” dimension is important to the research question because it describes the ways in which culture impacts how labour is distributed to Indian moderators based on their gender. Finally, as Guest, MacQueen, and Namey (2014) point out, using such formal mechanisms for theme identification is not always necessary: the “technical” dimension became a theme based on the narrative that had emerged regarding the need to train Indian moderators to override their cultural upbringing. As a result of analyzing the textual data in this way to classify themes (see Appendix C), three “dimensions” of content moderation in India were found and are discussed in the Analysis section of this report.

3.5. Ethics

This research was approved by the relevant ethics committee (reference number: SSH OII C1A 18 021. See Appendix A for CUREC approval). In terms of ethical considerations and interviewer bias in the context of this research, it is important to discuss the intercultural dynamics implicated in the interview process. One of the most critical ethical considerations taken into account was to avoid the pitfalls of ethnocentrism. Coming from a Western culture, special care was taken to avoid making value judgments about Indian culture, and to avoid evaluating whether the beliefs by which these moderators abide are too “traditional” or conservative compared to, for instance, Western liberal mores of sexuality. Additionally, as Latif (2002) notes, what may be appropriate in one culture may not be in another, and an awareness of forbidden topics and hidden taboos whilst interviewing interculturally is crucial. It was also important to be cognizant of an imbalance in cross-cultural power relations which could induce the respondents to reveal more information than they may have felt comfortable sharing (Shah, 2004). However, after reading into Indian customs and traditions and on

ruminating about a potential power imbalance, it did not appear as though there would be any issues that would compromise the ethicality of this research.

Given the very sensitive nature of content moderation as a profession in itself, preserving the anonymity and safety of the research participants is imperative. In order to protect the identities of the respondents, each interview quote in this research has been attributed to a pseudonym, and the names of each respondent's former and present employers have been redacted. Again, it was possible that the respondents would be divulging information in breach of an NDA, and as such, extensive measures like these were taken to mitigate any risks. In terms of ethical considerations in recognizing the emotional vulnerability participants may have potentially felt in discussing the darker side of content moderation, care was taken by avoiding overly salacious questions (Roberts, 2014), such as "what was the most graphic thing you have seen in your work?"

3.6. Limitations

In terms of limitations to the interview method used, there are special considerations with respect to reliability and validity that merit discussion. Reliability refers to the extent to which a study may be replicated; validity is concerned with the accuracy of scientific findings (LeCompte & Goetz, 1982). A shortcoming of this methodological approach in terms of internal validity is that interviewees may wish to withhold information for a myriad of reasons – being within direct violation of an NDA, fear of losing their jobs, or reliving the horrors of their work being chief among them. Though this is certainly a methodological concern worth addressing, the research participants were extremely forthcoming and, while acknowledging that moderation comes with exposure to disturbing imagery, did not seem perturbed.

It is critical to note that the "instrument" used in interviewing – i.e., the interviewer – also complicates the objectivity and reliability of the study. The biases of the researcher in qualitative study influence not only the interpretation of interviews, but other elements of the

methodology such as the coding schedule, the identification of certain themes, and even the conclusions that are drawn (Shenton, 2004). This is an inherent issue to external reliability since the same conclusions may not be drawn if a different individual were to follow even the exact same interview schedule and review the same transcription. However, reliability and validity can be bolstered by discussing explicitly and transparently any biases at the outset (Kuzmanić, 2009), as has been done in the Ethics section of this paper. Moreover, situating the researcher's own predispositions – which has been done throughout this thesis research – is a key criterion for strengthening confirmability and ensures that the findings are shaped by the respondent data and not the researcher's own motivations (Miles & Huberman, 1994).

Finally, there are methodological limitations to conducting interviews over Skype. While VoIP services flatten and nullify the geographic distances that would otherwise prohibit the research from being conducted (Iacono, Symonds, & Brown, 2016), the use of such a service could impact openness and thus credibility. In technology-mediated interviews, an additional concern is whether the interviewer loses the ability to notice nonverbal cues (Gibson & Hua, 2016). Interviewer inexperience may have been exacerbated by the inability to see certain visual signals over Skype, thereby perhaps moderately impacting the interpretation of the data but to an extent that is unlikely to be significant.

4. Findings

The interviews for this research lead to two sets of fascinating findings. First, they reveal the diversity and clientele for whom Indian moderators review content. It has been well established that larger platforms like YouTube and Facebook request moderation services, but who are the others? Second, the findings elucidate a typology of the various ways in which Indian moderators review content: in the press, moderators are cited as having to make decisions in only “just a few seconds” (Weber & Seetharaman, 2017). However, these findings demonstrate that Indian moderators engage in a combination of both rapid *and* slow analytical moderation.

4.1. Diversity of Projects and Clientele

The literature rather grievously fails to consider in any detail whatsoever how content moderation is performed for platforms other than the Internet “giants” like Facebook, Google, and YouTube. One of the most interesting findings of this research is the sheer diversity of the types of clients from all around the world that seek outsourced content moderation services from India. The research participants all serviced a range of clients from the brand-name e-commerce platforms like Amazon, to more niche websites such as Memedroid, a meme archive, and Camfrog, a video chat and instant messaging portal. Other clients unmentioned in the literature or popular press are Fiverr, an Israeli company for freelancing services, and a whole host of dating platforms including Positive Singles (catered to those with sexually transmitted diseases), Mingle2, OKCupid, Zoosk, twoo.com, and Mamba.

Each platform comes with its own specifications as to what should get moderated, and how. For example, in a document on “moderation models” provided by a research participant, one online dating website instructs moderators to pre-emptively (or in “real time”) remove a range of inappropriate imagery, including all “nudity images” such as a “male penis,” “female breasts, nipple, or vagina,” as well as sex toys; “hate” such as “ISIS flag for hate” and “Confederate flag for hate,” and violence such as “real or fake blood,” and “images where a violent act has already taken place, people are injured.” The client also bans “cartoons, drawings, anime” presumably because it wants users to represent themselves as they are. And, while some clients such as Mamba ask exclusively for the moderation of profile photos, companies like Zoosk require moderators to investigate profiles more comprehensively.

4.2. Nature of the Work: Rote Review Versus Analytical Thinking

Once again contrary to Western assumptions about the routinized clickwork of moderators, another one of the fascinating findings of this research that existing literature overlooks is the

fact that much of Indian content moderators' work goes beyond rote review of content to be deemed inappropriate or appropriate, but is instead a holistic and critical *analysis* of this content for accuracy and legitimacy. One of the criticisms against Indian content moderators is that they lack analytic ability, and as such are given content to moderate like pornography or spam that is more easily identifiable (see "Josh" in Roberts, 2014). However, not only are content moderators in India responsible for accepting or rejecting images that are in violation of a client's guidelines, but they also exercise a high degree of analytical thinking in content moderation by verifying the identities of individuals, their intentions, and more. Dating websites from around the world – across the United States, to Russia and Belgium – were the Indian moderation firms' primary clients. However, the duties of content moderators in India are greater than simply removing the horrifying imagery from these websites, though this is of course in itself a painstaking and labourious task. Rather, the moderators are additionally responsible for reading the dating profiles and ensuring that the information is true to help prevent vulnerable individuals from being exploited in malicious online dating schemes.

As the founder and CEO of one of India's largest content moderation firms, Sanjeet, puts it, the amount of porn and violence content moderators purportedly see is grossly exaggerated in the news – at his company, the primary objective is really to stop online fraud. Sanjeet says that his moderators work to stop these scams before they escalate too quickly: scammers may be real people with legitimate profiles but who try to manipulate an individual's emotional susceptibilities. Such actors with "evil intentions" will first pretend to be the "boyfriend or girlfriend" of a user on the site, and will then slowly begin to ask for small sums of money which gradually turns into greater and more urgent demands. It is a content moderator's job to understand the techniques used by these criminals to stop this from happening in the first place. A moderator who used to work at Sanjeet's firm described it as such:

Pritam: There was twoo.com. It's a Belgium product. It was really big and it was very interesting project. So it's a like – it's a dating site. There are so many scammers there. [...] We need to check first photos. From which location they are logging in. So we are having access to do everything. We can check their messages, from where they are

logging in, what they are talking, if they are giving anyone their numbers, we need to reject it.

It is worth noting that Pritam's explanation of his duties was memorably delivered with pride over the course of the interview. As will be discussed more thoroughly in the Analysis section of this paper, his description of the assignment as "interesting" runs counter to the dominant narrative about the work of content moderators, which is perpetually portrayed as harrowing and debilitating, lacking in even a modicum of enjoyment or personal fulfillment. Rather, Pritam takes his job very seriously, because he recognizes the harm that could come to users of the website if he does not monitor the contents of each profile stringently. Indian moderators think not only analytically about each profile they moderate, but the wider, reverberating effects of their work for others around the world. Another Indian content moderator, Vikesh, went into elaborate detail about the cues to look out for when moderating an online dating profile that could have potentially malicious intentions:

Vikesh: I'll show you some pictures. [...] They'll be mentioning their mail ID to contact personally on Zoosk. OK. If you see my mail ID there with gaps, in this way I cue you mail ID. [...] Even in this same thing, mobile numbers will be the same way. [...] They will give many hints to contact in some other parts.

Here, Vikesh is describing the process of authenticating and verifying a user's intentions. Vikesh describes the strategies that scammers will use to exploit users, primarily by encouraging contact off the platform through email or mobile. Since these platforms do not allow individuals to advertise contact details on their profiles, they will try to circumvent these rules by putting "gaps" in their email addresses ("mail ID" in Indian English) that automated systems cannot detect. Vikesh went on to say that he monitors for scammers who signal in a form of cryptography, where letters signifying numbers will be used to correspond to a mobile phone number.

Content analysis as opposed to purely the rote review of graphic content is performed for platforms and websites other than online dating. At his new place of work, where he is now a

team lead, Pritam's main content moderation client is Amazon. Here, Pritam is specifically responsible for once again not only ensuring that the graphic, nefarious user-generated spam is removed expediently from the platform before it even surfaces, but also for conducting quality assurance checks on the contents of the reviews that users leave behind. Working on Amazon video reviews, Pritam says that he takes ownership of matching what the users say about the product in the feedback video with the actual star rating they leave for the product, as well as ensuring the feedback videos are about the product the user has actually purchased. This takes on several different forms: for example, Pritam notes that if someone has bought a shirt, but the review discusses a pair of jeans, the video must be rejected. In circumstances where a product has been given a one-star rating, but the video review expresses satisfaction and describes the product in such a way so as to suggest it should be rated higher, the video will be rejected. This again demonstrates an impressive ability from the Indian moderators to infer nonverbal and verbal cues in a language that is not their mother tongue.

The literature has thus far focalized too myopically on only one aspect of content moderation, characterizing it as rapid-fire "clickwork". Rather, these findings suggest that *ex ante* content moderation is conducted on a spectrum; on one end of this spectrum are the faster, short bursts of moderation where the moderator reviews the content quickly to see if it is clearly vitriolic or an obvious violation of a company's guidelines. On the other end of this spectrum, however, is the more measured, calculated, and thoughtful element of content moderation, where a moderator must meticulously consider the intentions of the user, what that user has written, and then verify all of these details accordingly.

5. Analysis

After employing the methodological approach of thematic content analysis to the interview transcripts, three distinct themes materialized. The three broad themes discovered also directly challenge the assumptions made about Indian content moderators from policy experts and researchers in the West. The "structural" dimension demonstrates that while culture may not

have a direct impact on how content is moderated, as has been purported, it shows instead that Indian culture influences the way in which moderation work is distributed along asymmetrical gendered lines. The second dimension, the “technical” dimension, challenges the patronizing manner by which Western experts disparage an Indian moderator’s supposed inability to overcome his or her culture when applying content guidelines. Finally, the “subjective” dimension illustrates why generalizing the experience of American moderators to Indian moderators is erroneous: while US-based moderators remonstrate the low salaries and the indignity of the profession, Indian moderators by contrast see it as a stimulating career with high growth potential.

5.1. The Structural Dimension: The Gendered Division of Labour in India’s Content Moderation Industry

A revelation that emerged over the course of discussions with the Indian content moderators was the ways in which it became evident that certain types of content were moderated along gendered lines: according to both male colleagues and upper management, women were perceived as too sensitive and lacked the emotional constitution to be allowed to moderate the sexual and pornographic content that would oftentimes come across employees’ screens. This is made clear in the following anecdote:

Pritam: G-r-o-w-l-r. So that is project all about – it’s a gay site. Ok. So, when we are getting this kind of project we are not giving to any girls. Only nude images we are getting, only pornographics, only sex images we are getting, so we are not giving to any girls. When girls are coming, they are getting Snapdeal, in any marketplace sites they are getting, even... any whatever projects they are doing they are getting nude images, but very less images.

Here, one of the content moderators, Pritam, discusses a project he worked on for the company GROWLr. GROWLr is a mobile chat and social networking app for gay “bears,” a euphemism in the LGBTQ+ community for those who identify as heavy-set or muscular hairy, bisexual men. Pritam explained that the project received a regular onslaught of pornographic images, and explicitly states that this kind of imagery would not be appropriate for a female employee to moderate. Instead, he says, the female moderators would receive projects that were

considered safer, such as online marketplaces like Snapdeal, an Indian e-commerce website. Though there may still be nude photographs and sexually sensitive content to moderate on such online marketplaces, Pritam portends, there is significantly less of this content compared to that which he sees on GROWLr, making those clients more acceptable for women to moderate.

This bifurcation between what is considered a man's responsibility versus a woman's responsibility in the Indian content moderation industry is also mirrored in India's call center industry, particularly with respect to the range of tasks that are given to men and women based on stereotyped behavioural or cultural traits. For example, female employees in India are typically given client-facing or customer service roles because they are seen as more docile and amicable; contrastingly, male BPO call center employees are given responsibilities that include harassing credit collections from late-payers (Patel, 2010). As Nielsen and Waldrop (2014) note, BPO employment, and content moderation by extension, is viewed as a middle-class work environment: along with this comes the paternalistic views that the women who work in the industry need to be protected. This was corroborated in the conversations with another male Indian content moderator:

Kirit: To be very frank, in India we can't see that – men can be openly seeing everything, but women feel a little shy by seeing all those things, because at [former company] we have very bad project, [...] because it's social chatting, video call chatting. So there was much nudity. [...] We feel uncomfortable when lady watches and lady says, "What kind of work you are doing?" Suppose you came and see on my desk, I am watching for the nudity or the porn things, you can't just judge me over that.

In the above citation, Kirit reinforces what was previously articulated by Pritam: the amount of pornographic, sexual, and nude imagery that women in the content moderation industry are exposed to should be minimized or eliminated altogether where possible, but there is no cultural discrimination or stigma when it comes to men reviewing such content. More than this, Kirit claims that he himself feels discomfort when having to moderate sexually explicit content in the presence of women, and that he feels "judged" when a female content moderator glimpses what he is doing.

This difficulty reconciling male-female dynamics in the Indian content moderation space is perhaps what drove the creation and development of an all-male content moderation subsidiary in Hyderabad, which all the male content moderators interviewed had moved to after starting their careers in content moderation at larger firms in Bangalore. In discussions with the moderators, they cited that they were headhunted by this new firm, and were attracted by the opportunity to move to a new city in addition to the prospect of receiving a 40% salary increase on top of a 40% joining bonus. When pressed as to why the female content moderators from their former companies, which were gender-mixed, were not asked to join the new subsidiary, one content moderator said the following:

Pritam: When company started, in that place, there was not some good spaces to manage both boys and girls. So that is the reason we don't have any girls for now. [...] So that was a problem for my manager to handle *ki*, you know if any girls are coming they can't manage the nightshift and because it is very difficult to handle both boys and girls together.

Two elements of Pritam's explanation merit further analysis. The first is the implication that women cannot manage working the nightshift. There are multiple layers of connotation embedded in this statement: firstly, on the level of personal safety, incidences of female BPO workers getting raped or sexually assaulted on their journeys back home from working the nightshift are well documented (Patel, 2010). More than this, concern for young women working the night shift is less about the individual's personal safety, and more about how a young woman's presence in the urban nightscapes of India might affect a family's reputation (Nielsen & Waldrop, 2014). It is not exclusively about risking her safety against bodily violence, but also risking her social safety in the larger community. Thus, undergirding the inability to "manage" the nightshift is not only the physical threat such women might encounter, but damaging social ramifications as well. The second element of Pritam's comment that warrants deeper scrutiny is the observation that it is difficult to handle both "boys and girls" in the same space. As Kirit articulated previously, some of the male content moderators may feel judged for having to review inappropriate material in the presence of "shy" women.

Alternatively, this comment could be alluding to a long-standing criticism that is frequently directed against the BPO industry in India: that outsourced work like content moderation is morally corrupting the young Indians who work there, and that such work is wrong because boys and girls work together during the night and are exposed to Western ideas (Gopelwar, 2012). Thus, in addition to internal physical discomfort where proximity to female moderators may induce unease or embarrassment, there is an external, pervasive social stigmatization of mixed-gendered IT working spaces that involve virtually migrating between different cultural principles.

As Jyoti Puri (2002) writes, traditional markers like domesticity, marriage, and motherhood remain representations of what it is to be a woman in postcolonial India: middle- and upper-class women embody national cultural identity and are frequently subject to concerns of corruption particularly in the form of modernization and influence from the West. In sex education materials for teenagers in India, Puri (2002) found that girls and boys who “mixed freely” would garner a bad reputation due to “prevailing social mores,” and that this stigma would make marriage difficult in a country “where arranged marriages are still the norm” (p. 33). In India, the desire for boy children remains firmly entrenched, because they represent “social security” for parents while girls are expected to marry and tend to their new homes (Patel, 2010, p. 90).

These cultural and societal expectations that disproportionately affect Indian women, rather than men, have real and tangible economic consequences for women looking to advance their careers in the Indian IT sector. As was described above, the male content moderators who were interviewed predominantly chose to leave their mixed-gendered working environments due to the substantive salary increase offered by their new employer. However, because those working at this new content moderation subsidiary saw the physical and social barriers to women working the nightshift in a mixed-gendered office, as well as the fact that men were better equipped to moderate all forms of content where women were seen to be unable to handle the

sexual imagery, female moderators were never asked to join the new company and were hence barred from reaping the associated financial rewards. This is a point of cultural divergence in that men and women have completely different employability, precisely because Indian culture itself seeks to protect femininity. Women who work in the BPO industry in India often do not garner respect or acceptance despite earning relatively high wages: they are viewed with pity for having to work in the first place (Patel, 2010). This thus reveals the systemic stifling and denial of access to women based on discriminatory cultural norms that dictate women should not look at inappropriate content or stay out late, but rather, keep their place in the home. And since physical closeness between unmarried men and women in the nightscape continues to be strongly discouraged, this trend towards upward social mobility for men in content moderation while women are left behind shows no signs of being disrupted (Gopelwar, 2012).

Speaking to Sanjeet, who does employ both female and male content moderators, revealed that while some moderation firms like his happily employ women, there continue to be gendered nuances in the work that is given to them. When asked to comment about the male and female dynamics in the company, Sanjeet proudly exclaimed that the ratio of men to women was about 60/40 – and boasted that the “entire recruitment team was female.” Moreover, Sanjeet explained that while his moderators do not suffer anywhere near the degree of psychological trauma extolled in the literature, there are 24/7 hotlines and specialized “pink hotlines” specifically for his female moderators to contact in episodes of distress. Finally, Sanjeet used the metaphor of a battalion to articulate his perspective – where once armies only allowed male soldiers to fight, eventually women were allowed to join, too. Content moderation for female workers is similarly “just another job.” While at first glimpse Sanjeet’s views seem progressive, there remain underlying antiquated ideas of women in his comments that indicate the company’s policies are in need of revision. Firstly, while Sanjeet is proud of his company being an equal opportunity employer, an entirely-female HR team means that the 60/40 ratio previously stated includes fewer females in content moderation roles and reinforces

gendered notions of women's work in the IT sector – that is, being pushed out of getting to interact with the technology at all (Hicks, 2017). Secondly, while designated helplines for women is a step in the right direction, referring to these hotlines as “pink” hotlines is condescending at best, and misogynistic at worst.

Indeed, one of the crucial flaws of this sample size is its complete absence of the perspectives of female Indian content moderators. All those female Indian content moderators who were contacted in recruitment efforts declined to respond or did not agree to participate in this research, and snowball sampling proved ineffectual since the content moderators who had moved to a different company and agreed to participate found that all of their new colleagues, more than 200 employees in total, were men.

It has now been made abundantly clear that the belief that a woman's chastity be preserved is heavily entrenched in Indian society and is deeply culturally ingrained. In the context of content moderation practices in India, this insistence on shielding women from exposure to sexual imagery permeates the everyday business practices of Indian content moderation firms and results in the deliberate streamlining of content where the review of nudity and pornography becomes exclusively a man's task. Where originally it was alleged that culture influences the decision-making processes of Indian moderators, this research shows instead that Indian culture and societal perceptions of feminine fragility impact the ways in which content moderation work gets distributed by gender. This discriminatory division of labour based on culture ultimately prevents female moderators from achieving the same social and economic mobility as their male counterparts.

5.2. The Technical Dimension: Cultural Competence Training for Indian Content Moderators

Another theme that emerged throughout the course of conversations with the Indian content moderators, the Indian C-Suite executives, and the Silicon Valley-based policy managers, was the insistence on providing thorough cultural training. This dimension of analysis elucidates

that, contrary to the criticisms laid against Indian moderators, training is an important element of content moderation which the moderators themselves as well as Indian executive leadership take seriously. The data suggests that Indian moderators apply the guidelines methodically without the conscious injection of their own beliefs, despite how Trust and Safety managers claim their culture impacts the calibre of the moderators' work.

Klonick (2017) observes that while social networks like Facebook employ what they call "Community Standards" on a global level without localized or jurisdictional differentiation, all content moderators come "with their own cultural inclinations and biases" (p. 51). To ensure that these blanket community standards are applied in a manner that is consistent and uniform, it is necessary to minimize a content moderator's application of his or her own cultural norms and values (Klonick, 2017). One interviewee, Samantha, described precisely how difficult it is to craft globally applicable and translatable policies that are moderated the same when moderators come from different backgrounds. Before Samantha became a Trust and Safety lead at a large video streaming platform, she did much of the content moderation herself in conjunction with Indian moderators based abroad and noticed a disconnect on a decision with a video of a group of school boys fighting. To her, the video clearly violated the company's guidelines on minors engaging in violence, but the Indian moderators did not perceive it as such, since in their eyes, the minors were really adults. When asked about whether this misalignment in cultural perceptions and values between Indian and American moderators impacted accuracy and quality, she stated:

Samantha: Yeah, at some point we actually removed our India team altogether. It's a thing that I think, generally, like, a lot of companies face this issue – and it's truly just that we grow up with different sets of standards. And that's just so based on where you are. And it's not even standards, it's – what's the word I'm thinking of? It's just cultural norms, are just different. And so, like in India what would be considered sex, and sexy, and too like over the top, we're like, "But she's got her clothes on, what are you talking about?"

Samantha's strategy for alleviating the confusion that resulted from cultural differences was to make it "black and white" – if an Indian moderator saw X, it should always result in Y

action. In this way, it was not about “what was happening in the video” or “the meaning,” but rather a decision made based on the raw footage the moderator was seeing.

As a former legal removals lead at a large platform who is now a Trust and Safety program manager, Phil’s own experience echoes that of Samantha’s in having to teach the Indian content moderators how to bifurcate, on the one hand, what they were culturally taught should be considered appropriate, and on the other, what was culturally appropriate for the Western audiences for whom they were moderating. Phil was responsible for launching the operations team out of Hyderabad, India and flew to India to set up the office, hire the employees, and explain the tasks they would be working on. He related an issue in which the Indian team had difficulty grasping that much of the pornographic content that could be found through the platform was in fact legal, and that it should remain. As the platform is a search engine and merely a reflection of what can be found on the Web, links to websites with pornography appear in search results. If adult film companies find unauthorized copies of their content linked to in those search results, they may file a copyright claim and request the platform to remove the page from the results. If the notice is valid, the platform complies. For the first few weeks, the Indian specialists had difficulty applying the policies that went against their own intuition and cultural upbringing, removing the pornographic content that should have stayed:

Phil: I would say [it was] maybe like a cultural, I don’t want to say a cultural discrepancy, but like a cultural translation issue. There was a lot of porn that we had to review because one of the major requesters of copyright, of content removal on [platform name] is adult film companies. [...] So maybe there was some sensibility issues about what was appropriate and, you know, just because the content is inappropriate doesn’t mean that it needs to come down. [...] It’s not bad because it’s porn.

As Aneesh (2012) pithily puts it, however, “Training someone out of their culture and into another is never a simple task” (p. 522). And with all the rhetoric surrounding content moderation as custodial work in the process of “cleaning” the Internet, is it any wonder that an Indian moderator’s immediate inclination would be to delete a sexually dubious image? Klonick (2017) likens this process of training content moderators to overcome cultural biases

or emotional reactions to the content they review to the ways in which lawyers and judges are taught to apply rules to the facts of the cases they adjudicate. Ranging from “pattern recognition” to “professional judgement,” the same process is expressed by different names but ultimately refers to training content moderators through an iterative process to “override” cultural or emotional reactions and replace them with rational “valid” resolutions (p. 52).

Phil: I think once you start going through examples, and going through cases, and saying like, “Here’s an example of a time when you need to do this,” “Here’s when you need to escalate,” “This is what this looks like,” “Let’s think about applying that to future cases,” it started getting better. Through the years that we had this team it got much better.

This same process of teaching Indian IT workers to acculturate to the mannerisms and ways of thinking of those in the West is paralleled in the Indian call centre industry. For example, in his study of the BPO industry at an offshoring company called GoCom, Aneesh (2012) observed that cultural adjustments for Indian workers included having to learn American or British informal ways of speaking through formal training – agents did not have any prior knowledge of colloquial expressions such as “dude,” “jerk,” or “geek,” because they had not grown up hearing these Western slang phrases in India (p. 524).

However, in actually speaking with the moderators, it quickly became evident that the accusations of cultural misunderstandings were largely embellished. After asking one Indian moderator what he would do if he came across a bikini-clad woman on a client’s dating website, he answered in the following way:

Vikesh: When I am working for Mamba, which is a Russian website, I would follow Russian rules. [...] When I am working on UK website I should follow the UK rules. If the moderator is from India or else from some other area that is different, but rules for website is single.

Here, it is quite clear that Vikesh recognizes that despite his own cultural upbringing perhaps being different from the client for whom he is moderating, it doesn’t matter: he adheres to the guidelines and the “rules” of that client’s country, not his own. He claims it doesn’t matter whether the moderator is from India or elsewhere; the rules for the website is “single,”

meaning that it is not open to multiple interpretations but only the one that has been set out by the client.

In discussions with Aditya, the founder and COO of another Indian content moderation company, it is clear that management is aware that comprehensive training on the cultural mannerisms of their clients is necessary to moderate effectively.

Aditya: We have a strict three-weeks training before we deploy them onto any live project. [...] For example if you are working with any US clients, then our trainers will talk to those content moderators regarding the culture. And if you're talking about the you know, European countries, so each region has [a] specific kind of culture [...] As a moderator you have to know in and out of the countries, regions, technology, accent, everything, so only then can you do this as your job.

Aditya went on to explain that his company has been in the content moderation business now for over a decade, and the unique selling proposition he leverages against other companies is the fact that moderators are trained in-house. There are guidelines for every region and for every category – there are separate guidelines for video game clients, online dating website clients, and online publication or review sites. Then, on a more granular level, these top-level guidelines become region-specific.

Kirit, another Indian content moderator who previously worked for Aditya's company, had this to say about guideline enforcement:

Kirit: Yes, actually [we] get good guidelines from our clients. It's almost like 80 to 90 pages of guidelines. Our director is very good in these things because he is very particular about guidelines, so we get every day a quality report on basis of what we do. [...] Some of them are happy and some are not, because not every employee is as good as another, no?

Kirit notes that some quality reports are “happy” with the work of the moderators and sometimes they are not, because not all employees perform to the same high standard uniformly. His comment also strikes right at the very heart of what is so problematic regarding the literature and press on commercial content moderation in India thus far. When complaints about the quality of an Indian moderator's work are made, they are attributed to Indian *culture* and an inability to understand US-centric platform norms (see quote from Samantha, this section). When complaints as to the “inconsistent” application of guidelines are levelled against

content moderators in general, as in a Pro Publica piece on the uneven enforcement of hate speech rules on Facebook, they are attributed to “complex guidelines” (Tobin, Varner, & Angwin, 2017). This was affirmed by Sanjeet, who explained that in cases where Indian moderators make mistakes in applying policy guidelines, this may be due to a “selection error” in hiring the individual or a “training error,” but he vehemently disagrees with the assertion that it is a “cultural issue.” It is never considered that the difference in quality from Indian moderators is a result of individual bias, which moderators from *all* backgrounds and countries grapple with, but rather as a result of their culture and incomprehension of Western norms. The benefit of the doubt is not given to Indian moderators in the same way they are to moderators from elsewhere.

5.3. The Subjective Dimension: Unearthing Perceptions Amongst Indian Content Moderators

The final topic that emerged from a thematic analysis of the interview data was the subjective dimension of moderation for Indian content moderators. This section unearths the ways in which Indian content moderators understand the joys and difficulties of their work, as well as their general perceptions of content moderation. It was found that much of their experiences diverge remarkably from American moderators based in Silicon Valley, whose voices are the ones that dominate the literature on content moderation to date.

This section also problematizes the victimizing rhetoric surrounding the content moderation industry. That is, content moderation has been painted with a broad and generalizing brush, and is depicted as an industry wherein its workers suffer daily harms with no conceivable benefits or pleasure. The hyperbolic headlines all paint the industry as one that is wrought with toxicity: “Meet the Indian warriors who watch hours of beheadings, murders & gory content to clean the internet,” (Economic Times, 2017) “The Laborers Who Keep Dick Pics and Beheadings Out of Your Facebook Feed,” (Wired, 2014) and “The Worst Job in Technology: Staring at Human Depravity to Keep it Off Facebook” (The Wall Street Journal,

2017) are a select few among many. In one article from *The Guardian*, a former Facebook moderator even said that “there was literally nothing enjoyable about the job” and that “every day, every minute” he would see “heads being cut off” (Solon, 2017). As has already been mentioned, such depictions of the industry are crafted using contributions mainly from Western content moderators, with little thought or mention as to how these moderators living in India may be experiencing the work in a different manner altogether.

Notably, in direct contradiction to the sensational headlines underscoring the deleterious effects of content moderation, the Indian moderators interviewed for this research actually take pleasure in their work and see moderation as an exciting and viable career path. One of the comments that stood out prominently came from one moderator who proclaimed outright and happily that he enjoyed his work:

Pritam: I love my work so I don’t have any problem. Already I start my work doing video moderation, for Amazon project, so it’s good. [...] So I like [...] new projects like Amazon, Snapdeal, Flipkart, many good things at [company name]. We are getting different, different projects every two to three months. So definitely we are getting good ideas, new, new ideas... So that’s why still I am here.

Pritam pronounced more than once over the course of the interview that he enjoyed his work immensely. He did mention that there was a large degree of violence and gore that comes with having to moderate video content on Amazon, but that he genuinely came to enjoy his work after growing accustomed to seeing this imagery after the first few days. He now sincerely enjoys the work that he does because he has the ability to interact with the Internet on a variety of projects for large and prominent clients, while constantly learning about new ideas and topics.

Indeed, Sanjeet mentions that moderation being portrayed as watching porn all day is “just a myth” that journalists keep publishing, and notes that a positive attitude is pervasive among the employees at his content moderation firm in India:

Sanjeet: The thing is that they understand the importance of it, they don’t see it as an, I would say, as a very low-lying job. They know the criticality of their job. How critical it is. It is you know, something similar to your movie censors. So can we say a censor board, you know, if they are not satisfied with their job? No. That’s their job!

Phil mentioned how one way to keep morale up in the India office was to continuously show the moderators the positive impacts of their efforts, and highlighted how *TechDirt* once wrote an article about a piece of offending content they had caught, which he subsequently shared with the team. And while managers like Sanjeet and Phil instill in their content moderators an understanding of the importance of their work, the darker, more adverse side to this is the way in which complaints within certain companies are dismissed based on this premise. Aditya described the counselling services provided to content moderators at his company:

Aditya: Some employees have to face some of the content, especially when it is regarding the violence. Any suicidal stuff, attempt in the video. [...] End of the day, how we convince our employees, how we counsel our employees is, “This is our job.” As a moderator, we are here to safeguard our clients’ online reputations.

This is, frankly, a disturbing way to address complaints from employees. Rather than offering mental health programs to address psychological trauma, or even a help line or designated HR employee to listen, content moderators are effectively told to just “suck it up.” While it has already been established that mental health services for content moderators globally, even in North America, are deficient to non-existent, the fact that this type of counselling service is seen as appropriate or sufficient raises justified alarm. This corporate, client-oriented way of thinking that is systematically drilled into each employee also makes it difficult to distinguish whether the Indian content moderators interviewed genuinely enjoy doing their work, and are simply unfazed by the more disturbing content that they review, or if the need to do one’s job properly is so heavily impressed upon them that they have learned to accept the disturbing content as part and parcel of keeping themselves employed. When one of the moderators, Rahul, was asked how he felt about seeing disturbing content on GROWLr, he responded:

Rahul: That is our work, na? We have to moderate.

After asking Rahul whether the work makes him feel bad or uncomfortable, he responded “no”: the first few days are hard, but afterwards, “it will be ok.” Rahul’s dedication to his work demonstrates remarkable resilience on the part of Indian content moderators. And while at first glance it may seem that moderators are forced to enjoy their work, when examining precisely *why* the moderators claim they enjoy their jobs, the data suggests that they are sincerely excited to work on interesting projects and progress in what they consider to be a dynamic career. This is in direct contrast to the US-based moderators who quit their jobs within the first few days, or who even leave on their lunch breaks and never return (Weber & Seetharaman, 2017). When asked why he still continues being a content moderator despite having to sift through the Internet’s drudgery, Rahul responded:

Rahul: We got some opportunity from this company. [...] Opportunity – salaries, everything. Company facilities, company services. [...] I like working at [company]. Actually, I want to make this my professional.

Another moderator, Rishi, framed his enjoyment of the moderation industry in terms of “opportunities” the industry affords as well, and says he would recommend it to others:

Rishi: Yeah, it’s good. Superb. [...] Actually, this job, starting that day, I have been thinking, “How I can join this one?” But now no have any tension. Life is okay. Next ten years, twenty years, it’s no problem.

In the beginning, Rishi doubted he would be able to keep moderating, but after becoming acclimatized to the requisites of the job, he describes it as “superb” and foresees many opportunities. Rishi then went on to say that though the job is tough, “people manage.” This was echoed by another Indian moderator, Imran, who said he would also recommend a job in content moderation because he says it will become even “more popular” in the next five to six years than it is now. Thus, it would appear that the dominant narrative running through the writing on content moderation – that it is a depravity-ridden, depressing job – does not hold when speaking to a host of different moderators doing this work in India.

The stark juxtaposition between a moderator's experience in North America, and a moderator's experience in India, may also be due to the cultural and professional climate of each region. While in both regions young graduates are eager to start their careers, the opportunities for advancement, decent wages, safe working conditions, and legal protections differ drastically between them. In the Service Level Agreement (SLA) for one client, Indian moderators were expected to moderate ten *lakh* (one million)'s worth of images per day in an eight-hour nightshift from 10:00 p.m. to 6:00 a.m. In the U.S., a content moderator reviews about eight thousand posts per day and can be paid \$0.02 USD per image, this being on the lower end of the scale (Madrigal, 2017). Assuming that the Indian moderators made the same amount of money as their American counterparts, one million images at \$0.02 USD apiece would mean an average salary of \$20,000 USD per day. Yet, the Indian research respondents reported that they only made 8,000 INR to 13,000 INR per month at their first content moderation company (the equivalent of \$115 USD - \$200 USD) and up to 30,000 INR (\$430 USD) per month at their second company. The Indian content moderators review more images than their American counterparts for less money and with fewer complaints.

The repeated references to the "opportunities" in content moderation for Indians also warrant further analysis. As Casilli (2016) points out, in countries like the Philippines, Bangladesh, and India, technologically-mediated labour is presented as the best and only possible "future of work," and optimism surrounds these on-demand services and content farms to provide new opportunities (p. 22). Indians work around the clock, for meagre pay, without the ability to complain or provide feedback for fear of potentially losing their job, precisely because they are conditioned to believe a fruitful future is contingent on digital employment like moderation. The Indian BPO industry is also significantly more precarious than its Western equivalent – particularly when it comes to content moderation, decisions about Indian moderators' jobs are made on the other side of the world, without their input, and likely without consideration as to how such decisions would adversely impact the livelihoods of the Other.

The policies at these enormous global platforms as well as the labour critical to their efforts are overseen by “a few hundred, largely white, largely young, tech-savvy Californians who occupy a small and tight social and professional circle” (Gillespie, 2018, p. 130). As on-the-ground Indian moderators living in the Global South, these workers lack prestigious educations, affluence, and the social capital to influence the very same policy decisions that they implement day in and day out. As a Trust and Safety specialist working with an outsourced Indian content moderation team, Samantha noted that one of the biggest grievances she dealt with were the complaints from the moderators regarding their inability to influence the policies themselves:

Samantha: There’s – such a weird *angst* [emphasis in original] that gets built up between the people who are doing the review or answering the questions, and the people who are making the rules. And there are a lot of people in India who are like, “Well, we wanna make the rules!”, and “We wanna do this work!”, and “We’re watching this, we can do this.” But it’s this weird situation where this company is based in the US, and we need to have US-centric whatever it is, support or moderation.

Two elements of this quote immediately jut out as points for discussion. The first is the way in which Samantha seems indignant that the Indian moderators would have the audacity to want to craft the policies, characterizing this dissatisfaction as “angst” as though they should stay in their place at the bottom of the rung while upper-level executives in Silicon Valley get to make all of the decisions. This is indicative of a cycle of disempowerment that Indian content moderators are perpetually subjected to – doing what is considered the grunt work of moderation, without any opportunity to provide their input on the process based on their own experience or expertise. Indian content moderators should neither be seen nor heard. The second is the inherent contradiction in the need for US-centric policies – when companies claim they want to make their “community standards” as universally applicable as possible, this is simply untrue. They merely wish to make their policies, developed with American ideology and values in mind, as *palatable* as possible to what other countries might find appropriate. Kleiner (2016) refers to this as “digital colonization”: the predatory tactics that present-day platforms use to homogenize not only commodities, but also norms and standards. American conceptions of freedom of expression, nudity, hate speech, and violence will always take

precedence over other countries' views of these ideals, no matter how "globally translatable" these platforms endeavour to make their guidelines.

To summarize, there are widespread misconceptions about the content moderation industry resulting from the narrow way in which it is framed. CCM has been described both in the academic literature and popular press as a damaging and demoralizing profession; however, in speaking with Indian content moderators, it is evident that their subjective experience is dramatically different from moderators in North America. For them, CCM is an opportunity to make (what they consider to be) a decent wage, learn about topics and cultures from around the world, and fulfill a sense of purpose in keeping the Internet safe. They are imbued with a sense of pride that is in sharp contrast to the sanctimonious way in which American content moderators describe the job: moderation is work that many in the Philippines and India "aspire to," whereas "American moderators often fall into the job as a last resort" (Chen, 2014). In exposing the realities of Indian content moderators, this research decolonizes the work on content moderation to date.

6. Conclusion

In conclusion, this research amalgamates and analyzes interviews with three important groups working in content moderation to answer the research question, *In what ways do the cultural beliefs of Indian content moderators impact how content is moderated on online social networking sites?* While still a burgeoning academic field, this research rebuts many of the imperialist tones about the global CCM industry currently endorsed in the literature. As a result of globalization, content moderation is increasingly outsourced to countries such as India: in other studies featuring interviews with current and former American CCM workers, Indian moderators are accused of lacking the ability to surmount cultural biases when applying platform guidelines. American reviewers and policy specialists claim the differences in cultural connotations regarding topics such as sex and violence result in inferior levels of accuracy and quality from Indian CCM workers.

The data from this research suggests that these criticisms are overly simplistic. While Indian moderators are represented as simpleminded, they do much more than the frenetic keep-or-delete type of moderation by which CCM work is traditionally characterized. Rather, Indian moderators demonstrate an ability to review content in a way that is measured and calculated, analyzing the nuances behind images and posts all the while undergoing a complex process of acculturation to understand the global user communities of the clients for whom they moderate.

Thus, while culture is indeed implicated in the practices of CCM workers, it is not in the fashion that was originally anticipated. Instead of culture influencing the *decisions* of these moderators, culture impacts the structural, gender-based division of labour in a top-level manner. Next, Indian moderators undergo a highly sophisticated training process to override their cultural upbringing, where they learn about the conventions of their clients and are taught to apply their guidelines in a way that is tremendously technical. Finally, culture and social conditioning impact the subjective experiences of Indian content moderators, who should not be understood as victims and who perceive their work as enjoyable, rewarding, and necessary for career progression in a developing country where digital labour like moderation is the sole and definite future of work.

Based on the conclusions of this research, future study on the global CCM industry would do well to recalibrate current conceptions of both outsourced CCM workers and the practice of moderation in general. Firstly, to treat the perceptions of only a limited group of those in CCM as somehow universal is a mistake and a disservice to those who, as in the case of the Indian moderators, do the work under different conditions and as such possess different experiences. Next, claims against Indian moderators and the quality of their work being substandard due to “their culture” should be supplemented with empirical evidence, not just assumption. Lastly, content moderation is rapidly evolving into a necessary and global industry: to have any hope of portraying it accurately, all voices involved should be given an opportunity to speak, not just the ones that are traditionally privileged to begin with.

References

- Aneesh, A. (2012). Negotiating globalization: Men and women of India's call centers. *Journal of Social Issues*, 68(3), 514-533.
- Baker, S. E., & Edwards, R. (2012). How many qualitative interviews is enough?: Expert voices and early career reflections on sampling and cases in qualitative research. *National Center for Research Methods Review Paper*, 1-42.
- Braun, V., Clarke, V. & Rance, N. (2014) How to use thematic analysis with interview data. In *The Counselling & Psychotherapy Research Handbook* (pp. 183-197). London, UK: Sage.
- Casilli, A. (2016). *Is there a global digital labor culture?* Paper presented at the 2nd Symposium of the Project for Advanced Research in Global Communication (PARGC), Philadelphia, United States.
- Chen, A. (2014, October 23). The Laborers Who Keep Dick Pics and Beheadings Out of Your Facebook Feed. *Wired*. Retrieved from <https://www.wired.com/2014/10/content-moderation/>.
- Daniels, J. (2013). Race and racism in Internet studies: A review and critique. *New Media & Society*, 15(5), 695-719.
- Duarte, N., Llando, E., & Loup, A. (2017). Mixed Messages? The Limits of Automated Social Media Content Analysis. *Center for Democracy and Technology*.
- Feamster, N. (2018, March 21). Artificial Intelligence and the Future of Online Content Moderation. [Blog post]. Retrieved from <https://freedom-to-tinker.com/2018/03/21/artificial-intelligence-and-the-future-of-online-content-moderation/>.
- Foiwe Info Global Solutions. (2018). Text Moderation. *Foiwe*. Retrieved from <http://www.foiwe.com/text-moderation/>.
- Gibson, B., & Hua, Z. (2016). Interviews. In *Research Methods in Intercultural Communication: A Practical Guide* (pp. 5z-597o). Chichester, UK: John Wiley & Sons Inc.
- Gillespie, T. (2018). *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*. New Haven, CT: Yale University Press.
- Gopelwar, M. K. (2012). Global call center employees in India: Work and life between globalization and tradition (Doctoral dissertation, University of Bremen).
- Graham, M., Hogan, B., Straumann, R. K., & Medhat, A. (2014). Uneven Geographies of User-Generated Information: Patterns of Increasing Informational Poverty. *Annals of the Association of American Geographers*, 104(4), 746-764.
- Graham, M., Wood, A., Hjorth, I., & Lehdonvirta, V. (2016). *Digital Labour and Development: New Knowledge Economies or Digital Sweatshops*. Paper presented at the Digital Transformations of Work conference, Oxford, United Kingdom.

- Guest, G., MacQueen, M. K., & Namey, E. E. (2014). Themes and Codes. In *Applied Thematic Analysis* (pp. 49-78). Thousand Oaks, CA: Sage.
- Harwell, D. (2018, April 11). AI will solve Facebook's most vexing problems, Mark Zuckerberg says. Just don't ask when or how. *The Washington Post*. Retrieved from https://www.washingtonpost.com/news/the-switch/wp/2018/04/11/ai-will-solve-facebooks-most-vexing-problems-mark-zuckerberg-says-just-dont-ask-when-or-how/?noredirect=on&utm_term=.c1afc25aeabf.
- Hicks, M. (2017). *Programmed inequality: How Britain discarded women technologists and lost its edge in computing*. Cambridge, MA: MIT Press.
- Iacono, L. V., Symonds, P., & Brown, D. H. (2016). Skype as a tool for qualitative research interviews. *Sociological Research Online*, 21(2), 1-15.
- Infoesearch. (2018). Services: Content moderation. *Infoesearch*. Retrieved from <https://www.infoesearch.com/services/content-moderation/>.
- Kar, S., & Sarkhel, A. (2017, May 30). Meet the Indian warriors who watch hours of beheadings, murders & gory content to clean the internet. *Economic Times*. Retrieved from <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>.
- Kaye, D. (2018, June 18 – July 6). Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression. *United Nations, General Assembly*. A/HRC/38/35.
- Kleiner, D. (2016) *Mr. Peel Goes to Cyberspace. Resisting Digital Colonization*. Paper presented at Digital Bauhaus Summit 2016 - Luxury Communism, Neufert-Box & Deutsches Nationaltheater, Weimar.
- Klonick, K. (2017). The new governors: The people, rules, and processes governing online speech. *Harv. L. Rev.*, 131(1), 1598-1670.
- Kuzmanić, M. (2009). Validity in qualitative research: Interview and the appearance of truth through dialogue. *Horizons of Psychology*, 18(2), 39-50.
- Latif, S. (2002) A theoretical framework for analysing cross-cultural perspectives. *Quarterly Journal of Gender and Social Issues*, 1(1), 63-74.
- LeCompte, M. D., & Goetz, J. P. (1982). Problems of reliability and validity in ethnographic research. *Review of educational research*, 52(1), 31-60.
- Lewis, R. D. (2006). *When cultures collide*. Boston, MA: Nicholas Brealey Publishing.
- Madrigal, A. C. (2017, December 15). The Basic Grossness of Humans. *The Atlantic*. Retrieved from <https://www.theatlantic.com/technology/archive/2017/12/the-basic-grossness-of-humans/548330/>.

- Miles, M. B. & Huberman, A. M. (1994). *Qualitative data analysis: A methods sourcebook*. (3rd edition). Thousand Oaks, CA: Sage.
- Nakamura, L. (2016). *Workers Without Bodies: Digital Labor, Race, and Gender*. Paper presented at the Terms of Media II: Actions conference, Cogut Center for the Humanities, Rhode Island, United States.
- Nielsen, K. B., & Waldrop, A. (2014). *Women, gender and everyday social transformation in India*. London, UK: Anthem Press.
- Patel, R. (2010). *Working the night shift: Women in India's call center industry*. Stanford, CA: Stanford University Press.
- Puri, J. (2002). *Woman, body, desire in post-colonial India: Narratives of gender and sexuality*. New York, NY: Routledge.
- Roberts, S. T. (2014). Behind the screen: The hidden digital labor of commercial content moderators (Doctoral dissertation, University of Illinois at Urbana-Champaign).
- Roberts, S. T. (2016). Digital refuse: Canadian garbage, commercial content moderation and the global circulation of social media's waste. *Media Studies Publications*, Paper 14.
- Roberts, S. T. (2017). Content moderation. In *Encyclopedia of Big Data*. UCLA.
- Roberts, S. T. (2018). Digital detritus: 'Error' and the logic of opacity in social media content moderation. *First Monday*, 23(3).
- Ryan, G., & Bernard, H. (2003). Techniques to identify themes. *Field Methods*, 15(1), 85–109.
- Saldaña, J. (2009). *The coding manual for qualitative researchers*. Thousand Oaks, CA: Sage.
- Seidman, I. (2006). *Interviewing as qualitative research: A guide for researchers in education and the social sciences* (3rd ed.). New York, NY: Teachers College Press.
- Shah, S. (2004). The researcher/interviewer in intercultural context: A social intruder!. *British Educational Research Journal*, 30(4), 549-575.
- Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information*, 22(2), 63-75.
- Solon, O. (2017, May 25). Underpaid and overburdened: The life of a Facebook moderator. *The Guardian*. Retrieved from <https://www.theguardian.com/news/2017/may/25/facebook-moderator-underpaid-overburdened-extreme-content>.
- Tobin, A., Varner, M., & Angwin, J. (2017, December 28). *Facebook's Uneven Enforcement of Hate Speech Rules Allows Vile Posts to Stay Up*. ProPublica. Retrieved from <https://www.propublica.org/article/facebook-enforcement-hate-speech-rules-mistakes>.
- Weber, L., & Seetharaman, D. (2017, December 27). The Worst Job in Technology: Staring at Human Depravity to Keep It Off Facebook. *The Wall Street Journal*. Retrieved from

<https://www.wsj.com/articles/the-worst-job-in-technology-staring-at-human-depravity-to-keep-it-off-facebook-1514398398>.

West, S. M. (2017). Raging against the machine: Network gatekeeping and collective action on social media platforms. *Media and Communication*, 5(3), 28-36.