

## LSE Research Online

#### Jacob Breslow

## Moderating the 'worst of humanity': sexuality, witnessing, and the digital life of coloniality

### **Article (Accepted version)** (Refereed)

#### **Original citation:**

Breslow, Jacob (2018) Moderating the 'worst of humanity': sexuality, witnessing, and the digital life of coloniality. Porn Studies. ISSN 2326-8743 (In Press)

© 2018 Informa UK Limited

This version available at: <a href="http://eprints.lse.ac.uk/87665/">http://eprints.lse.ac.uk/87665/</a>

Available in LSE Research Online: April 2018

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LSE Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. You may freely distribute the URL (http://eprints.lse.ac.uk) of the LSE Research Online website.

This document is the author's final accepted version of the journal article. There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

# Moderating the 'Worst of Humanity': Sexuality, Witnessing, and the Digital Life of Coloniality

An estimated one hundred thousand people worldwide work as content moderators, responding to the millions of photos and videos uploaded online every minute. Primarily employed by outsourcing companies in the Philippines, these labourers scrub social media of sexual content. This article unpacks what it calls the 'digital life of coloniality' as it is produced through content moderation along two lines of interrogation. It initially suggests that the traditional understanding of the coming together of sexuality, subjectivity, and regulation under colonialism are rendered more complex by content moderation, positioning the formerly colonized as regulators of their former colonizers' sexualities. Secondly, asking questions of witnessing, ethics, and accountability, it interrogates the lines of disavowal and displacement which structure the offshoring of violent, obscene, and mundane sexual content. Contributing to the field of porn studies, it suggests that the ambivalent and multiple directions of sexual subject production within digital coloniality be addressed anew.

Keywords: coloniality; content moderation; pornography; witnessing; social media; outsourcing

#### Introduction

While official numbers are difficult to ascertain, there are, according to some reports, over one hundred thousand people worldwide whose job it is to moderate the hundreds of millions of photos and videos uploaded, and statuses updated, every minute to various social media sites. Companies such as Facebook, Twitter, YouTube, and Instagram, all rely on what Adrian Chen has called 'an army of workers employed to soak up the worst of humanity in order to protect the rest of us' (2014). This work, what Sarah Roberts (2016) calls 'commercial content moderation', ranges in scale from the reading, filtering, and deleting of written speech in comments and statuses (a process which, particularly in regards to hate speech, has its own difficulties), to the viewing, flagging, and deleting of photos and videos of violence and sexual content. While new

technologies are being developed to do some of this work through algorithms and 'photoDNA', ii the majority of this labor is undertaken by people employed by outsourcing companies in the Global South. Due to different companies having different standards for what is or is not acceptable, and due to these companies often relying on users to be the first line of call (reporting already uploaded content they deem inappropriate), flagged sexual content can include everything from depictions of 'female nipples' and bodily fluids, to bestiality, child pornography, mundane sexual expression, art, rape, and sexual torture. While some of this work is done in the United States, often by temporary workers or low-paid sub-contractors, most of these companies, Catherine Buni and Soraya Chemaly report, 'consign their moderators to the margins, shipping their platforms' digital waste to "special economic zones" in the Global South' (2016). Buni and Chemaly's use of the term 'waste' is an indicative one. Indeed, the lines of movement from production to disposal of online content are often the same ones, Roberts argues, which direct the flow of physical waste (and particularly e-waste) as well (2016). These digital lines of movement—facilitated by and made physical through deep-sea cables that stretch around the globe—directly follow, Roberts writes, the 'well-worn circuits established during periods of formal colonial domination and continuing now, via mechanisms and processes that reify those circuits through economic, rather than political or military, means' (2016, 6). In other words, the circuits that transport the gamut of sexual content flagged for moderation in the United States and elsewhere to workers primarily based in India and the Philippines, are part of the contemporary digital life of coloniality, a coloniality which is, as has always been the case, structured through and productive of sexuality.

In this article, I work to unpack the digital life of coloniality as it is produced through commercial content moderation along two lines of interrogation, both of which

are specifically focused on the ways in which sexuality is structuring postcolonial relations in this particular contemporary moment. My first contention is that under the digital life of coloniality, the traditional understanding of the coming together of sexuality and regulation under colonialism and coloniality is rendered far more complex. What does it mean, I ask, that the contemporary regulation of 'good' and 'bad' sexual content—and particularly of the content which is specifically available for users in the United States—is being undertaken by workers in the Philippines, a former colony?<sup>iii</sup> Does the fact that this regulation of American sexual content is done by people in a former US colony, rather than vice versa, open up the ways in which sexuality must be understood in relation to colonial power? In asking this question, however, my analysis necessarily turns to a second line of explication. For my concern is not just, or not really, the issue of regulation (particularly as this question of the direction of regulation between the United States and the Philippines is, and has always been, rather complex). My analysis of the digital life of coloniality lies rather at the level of affect and accountability. Here, in the second part of this paper, I shift from the question of regulation to that of witnessing, and I ask after the colonial lines of disavowal and displacement which structure the offshoring of commercial content moderation. For the offshoring of this labor is not just, I shall argue, a question of political economy. It is rather a means of displacing the affective, ethical, and political act of witnessing—and thus the questions of accountability which accompany this act to former colonies as well. Across this article, then, I contribute to the critical field of porn studies by suggesting that the ambivalent and multiple directions of sexual subject production within coloniality be interrogated anew in our scholarship. Doing so, I offer up a postcolonial critique with a focus on the digital to trouble the geopolitical

locatedness of the longstanding debates within and beyond feminism about obscenity, pornography, and censorship.

#### **Close Cultural Ties**

In his 2014 article for *Wired*, titled "The Laborers Who Keep Dick Pics and Beheadings Out of Your Facebook Feed", Chen documents what he calls a 'vast, invisible pool of human labor' which is employed to scrub social media sites. Giving an account of the everyday labor of Baybayan, an employee for TaskUs, an outsourcing company which operates globally and which employs workers like Baybayan to moderate the app Whisper, Chen writes:

Watching Baybayan's work makes terrifyingly clear the amount of labor that goes into [moderating] Whisper ... He begins with a grid of posts, each of which is a rectangular photo, many with bold text overlays ... A list of categories, scrawled on a whiteboard, reminds the workers of what they're hunting for: pornography, gore, minors, sexual solicitation, sexual body parts/images, racism. When Baybayan sees a potential violation, he drills in on it to confirm, then sends it away—erasing it from the user's account and the service altogether—and moves back to the grid. Within 25 minutes, Baybayan has eliminated an impressive variety of dick pics, thong shots, exotic objects inserted into bodies, hateful taunts, and requests for oral sex. (2014)

These laborers, Chen notes, are employed across a number of locations across the globe, but, like Baybayan, they are predominantly based in the Philippines.<sup>iv</sup> This is so, Chen writes, because the Philippines, as a former US colony, has 'maintained close cultural ties to the United States, which content moderation companies say helps Filipinos determine what Americans find offensive' (2014). Speaking about the 'upsurge' of outsourcing in the Philippines, an upsurge which recently saw the Philippines surpass India as the world leading site of outsourced labor, Emmanuel David writes that this shift to the Philippines 'has been shaped, in part, by the country's long history as a US

colony. It boasts a sizable English-speaking population, one with a vexed postcolonial relation to American culture' (2016, 382). As David and others note, it is this particular postcolonial relation to American culture that is often used as justification for the bulk of this labor being done for low-cost in the Philippines (cf. Isaac 2017). Content moderators in the Philippines often make in a day what some US based moderators will make in an hour.

The colonial connection between the US and the Philippines is an important one, employers argue, because the task of content moderation requires a particular understood as exceptionally American—ethical sensibility. Much of this sensibility is, as one might expect, wrapped up in heteronormativity, familial respectability, reproductive futurity, and Catholic moralizing. As Buni and Chemaly write, for example, YouTube's Safety, Quality, and User Advocacy Department (SQUAD) assess content through the following question: 'Can I share this video with my family?' (2016). Clearly, a particular familial respectability is at play here, as numerous reports have come out over the last couple years about YouTube (as well as other companies) flagging and restricting LGBT content in a bid to make their sites more 'kid friendly' (Chokshi 2017). And yet, the forms of desire and sexuality that are being regulated through content moderation do not so easily fit within a heteronormative, or even homophobic frame. According to Emma Barnett and Iain Hollingshead, Facebook has a 'fascinatingly strict set of guidelines determining what should be deleted' (2012). Sounding like a convoluted version of Gayle Rubin's charmed circle (1984), these guidelines stipulate that:

Pictures of naked private parts, [...] and sexual activity (apart from foreplay) are all banned. Male nipples are OK, but naked breastfeeding is not. Photographs of bodily fluids (except semen) are allowed, but not if a human being is also shown. (Barnett and Hollingshead 2012)

According to Facebook's own 'community standards' (as of June 2017), 'photographs of people displaying genitals or focusing in on fully exposed buttocks' are regularly removed, and images of 'female breasts' are restricted 'if they include the nipple', but an exception is now made for photos of women 'actively engaged in breastfeeding or showing breasts with post-mastectomy scarring.' Here, the ambivalence surrounding the nipple and the sexed body it is attached to, activities it is engaged in, and surgery it has undergone, as well as the rapid shifts in the distinction of 'good' and 'bad' content between 2012 and 2017, make clear the hyper-responsiveness and vacillation of the industry and the difficulties of this type of labor. And yet, it is apparently these particular 'American' sexual and cultural values that workers in the Philippines share, making the placement of this work there not just economical, but 'logical' too.

In one of the more interesting accounts of the decision making process behind the moderating of images, Buni and Chemaly discuss their interview with Charlotte Willner, the Safety Manager at Pinterest, wherein she recounts the ways in which moderation policies had to shift after the success of *Fifty Shades of Grey*:

[Charlotte] and her team were hustling to develop new BDSM standards. 'We realized,' she later explained by email, 'that we were going to need to figure out standards for rape and kidnapping fantasy content, which we hadn't seen a lot of but we began to see in connection with the general BDSM influx.' The calls were not easy, but it was clear that her team was making decisions on a remarkably granular level. One user was posting fetish comments about cooking Barbie-size women in a stew pot. Should this be allowed? Why not? (2016).

While it was agreed that the images could remain—they were deemed implausible as a practical threat, as 'a full-size woman can't fit into stew pot, the team figured' (ibid.)—I highlight this example because it is a clear indication of the type of ethical decision making, and sexual moralizing, that content moderators do on a daily basis. vi And while

Charlotte is based out of the San Francisco Bay Area, the policies that her team (and teams like hers at other companies) create, become standard practice for workers in the Philippines.

What I am most interested in exploring here, is how the colonial relation allows for, or justifies, the particular type of work that these content moderators do to be 'best placed' in the Philippines. Why, in relation to this latter example of Pinterest, for example, is the initial ethical call made in the US, while the everyday labor of enforcing this decision is facilitated by outsourced labor? What, specifically, does this outsourcing tell us about how sexual regulation is facilitated by, and facilitates, the digital life of coloniality? For indeed, part of what is interesting here is that the outsourcing of this labor clearly shifts who is traditionally understood as undertaking the work and burden of policing—blocking, censoring, flagging, erasing, and blurring—sexuality. Returning to Rubin's seminal "Thinking Sex" from 1984, it is clear that the sites of institutional and interpersonal power which defined and policed erotic hysteria in the US (ranging from the Supreme Court, to the field of psychology, to popular culture and the family) now also need to include technology companies and the outsourced workers who undertake the quotidian labor of moderating digital expression. Rubin's analysis, which mostly ignores the imbrication of sex panic with racialization and colonialism (perhaps a product of an affinity for Foucault), is still a useful framework for thinking through the effects of content moderation. Yet, in order to have a more in-depth understanding of what is at stake here, our analysis cannot take the 'nation' as a coherent, sealed-off entity in the same way Rubin does, particularly if this means not including outsourced laborers as central to the techniques of sexual surveillance and prohibition. Following Victor Román Mendoza's call to interrogate the 'metroimperial' fantasy of the US 'nation' as a bounded, bordered entity not actively engaged in imperial (dis)possession,

I want to consider this outsourced labor as fundamental to, and co-constitutive of, the regulation of sexual moralism within and beyond the space of the 'nation'. Indeed Mendoza, who is interested in the colonial relation between the US and the Philippines around the turn of the twentieth century, argues that the US needs to be understood as a metroimperial project produced through 'synchronic legal, material, ideological, cultural, and social exchanges across transpacific space' (2015, 9). This framing opens up Rubin's intervention to transpacific processes of regulation.

Doing so requires an analysis of colonial sexual regulation that departs from much of the wider scholarship on sexuality and coloniality. While there has, of course, been a history of scholarship which has discussed the multiple, hybrid, and contested ways in which sexuality, gender, and race have structured the subject production of the colonizer under colonialism, and in the wake of coloniality (Bhabha 1991; Fanon 1967; Spivak 1988; Stoler 1995; Woollacott 1999), this subject production is most often understood as emerging from the regulations which were specifically directed at the colonized. For example, Monique Mulholland's insightful analysis of the 2007 restrictions of access to pornography in many of Australia's Aboriginal communities in the Northern Territory, points out the ways in which the uneven enactment of zones of pornography prohibition (under the guise of child protection) functioned as a colonizing technique of control and pathologization (2016). As Mulholland argues, Mulholland writes:

In this case, porn functions (as it has always done) to entrench the degeneracy of abjected raced others. When 'porn' and 'Indigenous' are put together [...] a picture emerges of an inherent, internal pathology [which requires] intervention from a civilizing hand. (2016, 45)

In Mulholland's example, the restrictions on pornography, which were only enforced in Aboriginal communities, were precisely about controlling and regulating displaced indigenous communities.

Similarly, in M. Jacqui Alexander's analysis of sexual citizenship in Trinidad and Tobago (1994), the state policing of sexuality (and particularly non-reproductive and lesbian and gay sex) is also an effect of colonization, albeit under different terms. In her analysis, that is, the anxieties about the sovereignty and salience of the newly postcolonial independent republic led Trinidad and Tobago to reassert national sovereignty through an increased regulation of non-heteronormative sexuality. The implementation of the 1986 Sexual Offences Act, Alexander writes, allowed the postcolonial state to 'consolidate the myth of imperial authority' through regulating sexuality, and specifically through naturalizing heterosexuality within law (1994, 8). I highlight these two examples both because they are particularly compelling analyses of the workings of sexuality under conditions of coloniality, and because they are indicative of the way in which scholarship has come to understand the orientation of power in relation to colonialism (see also: Levine 2000; McClintock 1995; Povinelli 1997; Puri 2014). For these scholars, the colonial and postcolonial life of sexual subjectivity directs restriction, violence, stigma and othering at currently and formerly colonized populations.

What I am arguing in relationship to content moderation, however, is that the sexual regulation which is taking place under the digital life of coloniality, while still being facilitated through, a tactic of, and justified under, colonial and postcolonial power, is a regulation that is specifically directed at the sexualities of the former colonizers. Unpacking this statement is not to argue for a re-centring of the colonizer, nor to argue that Americans are somehow the victims of their colonizing of the

Philippines. Nor is it to suggest that content moderation does not also follow traditional lines of power and sexual subjection. Indeed, one of the complex effects of content moderation is the ambivalent production of a very particular range of sexualities for the workers in the Philippines. Chen, discussing the psychological effects of content moderation, relays this in the following account of his conversation with Denise, a psychologist who consults for two different content moderation companies in the Philippines. Speaking to the issue of high employee turnover, he writes:

Workers quit because they feel desensitized by the hours of pornography they watch each day and no longer want to be intimate with their spouses. Others report a supercharged sex drive. 'How would you feel watching pornography for eight hours a day, every day?' Denise says. 'How long can you take that?' (2014).

The act of moderating, in other words, is in and of itself a sexual one. It produces a complicated sexual subjectivity. As this above quote makes clear, workers undertaking this labour become 'desensitized,' their intimate and affective bonds with their partners become ruptured, or, alternatively, intensified. The very act of witnessing this material becomes ingrained within the psyches and subjectivities of these workers. As I shall argue in the latter half of this article, regulating digital space is thus an act of curtailing an exposure to sexual imagery in one place, while producing a hyper-sexualized environment in another.

While this sexual subjectivity may not be premised on, or productive of, an outright pathologizing of 'Filipino sexuality', it does, nonetheless, clearly produce a perhaps ambivalent libidinal effect. It also follows, as I argue via Mendoza just below, a longer historical co-production of the hyper-sexualized racial-sexual Filipino subject and the queerness of American desires. In this way, content moderation is as much a reversal of the usual framework for understanding the direction of sexual colonial regulation, as it is the production of an ambivalent sexual subjectivity in the former

colony. As such, my insistence on interrogating the multiple orientations of this regulation is precisely because there are continued forms of economic, psychic, and social relations of dependency, resistance, and elicitation which are facilitated through the quotidian acts of regulating the American sexual appetite. My opening up the digital life of coloniality to interruption and interrogation, is thus a call for the scholarship on sexuality and coloniality to continue to unpack the multi-directionality of regulation under coloniality.

The scholarship which has already undertaken this line of interrogation provides a useful framework for thinking this through. Mendoza, for example, argues that 'the kinds of intimate and even perverse relations between the figure of the Philippine subject and other people that emerge[d]' under American colonization in the Philippines, were 'not peripheral or contrary to the hetero-masculinizing, genocidal project of U.S. imperialism but constitutive of it' (2015, 2). Mendoza's book, which covers a vast amount of material, and includes an impressive array of archival research, makes many important arguments about how the American-Philippine relationship was structured through different forms of racial-sexual governance. In it he argues that while the racial-sexual regulation being undertaken in the Philippines was both criminalizing of the always-already 'queer' Filipino subject (27-30), and was also encouraging forms of hetero-marital relations between American soldiers and Filipino women, there was also a simultaneous opposite direction of regulation undertaken by the American colonial officers which specifically regulated and pathologized American sexuality.

To give just one example from his text, Mendoza documents how the American colonial administration of the Philippines specifically targeted and pathologized American sexuality through the close regulation of the sexual health of American soldiers. While the typical colonial and Orientalist discourse of the tropical environment

of the Philippines inciting 'dangerous erotic impulses' (2015, 55) in otherwise innocent American soldiers was operating at the time, it was additionally compounded, Mendoza argues, by a concern over the questionable sexual morals of the American soldiers who wished to enlist their services in the Philippines. Making this point, Mendoza cites a 1901 study by Captain Edward Munson of the US Army's Medical Department, titled The Theory and Practice of Military Hygiene. Munson 'cautions against admitting men whose criminal records show "conviction of felony," "masturbation," or "sodomy," writing, "the moral character [of the recruits] should be scrutinized with care in order that enlistments from the vagrant and criminal classes may be avoided" (2015, 55). Part of the scrutiny that was undertaken was a routine genital inspection for prospective recruits, and A central discursive consequence of the policing of American soldiers' sexual health was the production of a wider framing of the American populace as having questionable sexual morals, and poor sexual hygiene prior to their enlistment. While this regulatory orientation should fundamentally be understood as a lack of concern for the sexual health and wellbeing of Filipinos under colonial administration, it is important not to understate the implications of this discourse of American sexual deviance, which was produced through the act of colonization.

For Mendoza, then, the racial-sexual production which was central to the colonial administration of the Philippines pathologized, criminalized, and co-constituted both Filipino and American sexualities. This transpacific stigmatization of queer and non-normative sexualities differs, in that sense, from other accounts of the multi-dimensionality of colonial sexual regulation which frame the orienting of power's normalizing grasp towards the colonizers as a form of bourgeois, racial-sexual *cultivation* instead. In the most notable example of this, Ann Laura Stoler argues that, on one hand, 'the regulatory mechanisms of the [European] colonial state were directed

not only at the colonized, but as forcefully at "internal enemies" within the heterogeneous population that comprised the category of Europeans themselves' (1995, 96). On the other hand, Stoler argues, these regulatory mechanisms at play within Europe established 'new interventions in the governing of the self [which] legitimate[d] increasing intervention in the ethics of conduct, geared to the management of 'how to live'' (96-97). For Stoler, the 'management and knowledge of home environments, childrearing practices, and sexual arrangements of European colonials' (97) was precisely a project of the cultivation of the European bourgeois body. Both Mendoza and Stoler, then, argue that understanding the imbrication of sexuality and coloniality requires thinking about the multiple directions of regulation and subject production which are central to sustaining and proliferating colonial logics.

Arguably, the labor of content moderation should also be understood as a project of normative sexual cultivation facilitated by coloniality. The colonial logics of 'protection', expelling 'waste', and establishing an 'ethics' of healthy digital conduct underscore the biopolitical civilizing project of this labor. Chen's statement, cited above, that content moderators 'soak up the worst of humanity in order to protect the rest of us' (2014) works exactly within this vein, as it clearly positions the 'us' of this arrangement as a particularly located population deserving of, and benefitting from, the precarious labor of distantly placed others. Here, then, the 'close cultural ties' that bind the US and the Philippines are the same conditions which justify the separation of 'us' from 'them' (cf. Mirchandani 2012). The discursive construction of an 'almost the same but not quite' (Bhabha 1991, 122) Filipino racial-sexual subject, produces an ambivalent mimesis premised on exploitation, distance, and invisibilization.

In this vein, what differentiates Stoler's context and my own—besides the obvious differences in geopolitical context and temporal location—is indeed the

invisibility and effacement of the regulation which is being undertaken in the contemporary digital life of coloniality. Unlike the authors and publishers who produced, wrote, and distributed colonial household manuals across Europe, for example, the laborers scrubbing the internet of certain sexual content are relatively unknown and invisibilized. Speaking about the concealing of this labor, Roberts suggests that 'If there's not an explicit campaign to hide it, there's certainly a tacit one. [...] It goes to our misunderstandings about the Internet and our view of technology as being somehow magically not human' (Roberts, cited in Chen 2014). Similarly, Stone writes:

Internet companies are reluctant to discuss the particulars of content moderation, since they would rather not draw attention to the unpleasantness that their sites can attract. [...] Outsourcing companies are also reluctant to discuss the business on the record, since their clients demand confidentiality. (2010)

This effacement is important, I argue, because it indicates that more is taking place in the offshoring of this labor than simply the 'necessary' regulation of sexual content, and the financial incentives for technology companies. More is at stake, that is, than the misconception that the internet is a magical posthuman zone of possibility. There is, as I shall show in the following section, a fundamental difference between how content moderators approach sexual content and other political content (often by allowing the 'properly political' material to remain available for users while erasing any and all sexual content). And it is because of this, that the invisibility of this labor seems to function less in relation to the myth of the internet, and more as a disavowal of the very humanness of sexual violence and a displacement of the traumatic psychological effects of 'cleaning up' humanity's digital footprint.

#### The Displacement of Witnessing

It is precisely this line of questioning which leads me to the analysis that forms the rest of this article: that of witnessing and accountability. In what remains of this article, I bring together my above discussion on the sexual regulation at play within coloniality's digital life with the scholarship on witnessing to explore what is at stake in the outsourcing of content moderation along the lines of trauma, affect, and accountability. Doing so requires thinking across various scales and effects of traumas, and it entails risking a slippage into abolitionist perspectives on pornography, and into a sex negativity, which I do not subscribe to. Despite these risks, which I discuss further below, I undertake this line of questioning because I am interested in parsing out the psychic and political work that offshoring allows in relation to witnessing and being accountable for the 'worst' of humanity.

As a way of making more clear this shift from the multi-directionality of regulation under the digital life of coloniality to the questions of affect, trauma, and witnessing which I am arguing are absolutely entangled with the outsourcing of content moderation, let me return to Babayan, the labourer I introduced at the start of the article. Describing his first encounter with Babayan, Chen writes:

The campuses of the tech industry are famous for their lavish cafeterias, cushy shuttles, and on-site laundry services. But on a muggy February afternoon, some of these companies' most important work is being done 7,000 miles away, on the second floor of a former elementary school at the end of a row of auto mechanics' stalls in Bacoor, a gritty Filipino town 13 miles southwest of Manila. ... In a large room packed with workers manning PCs on long tables, I meet Michael Baybayan, an enthusiastic 21-year-old with a jaunty pouf of reddish-brown hair. If the space does not resemble a typical startup's office, the image on Baybayan's screen does not resemble typical startup work: It appears to show a super-close-up photo of a two-pronged dildo wedged in a vagina. I say *appears* because I can barely begin to make sense of the image, a baseball-card-sized abstraction of flesh and translucent pink plastic, before he disappears it with a casual flick of his mouse. (2014)

As I argued above, this is labour that is clearly intended to regulate the borders of sexuality online. It is labour, as Chen writes, that social media companies have depended on in order to grow into a multibillion-dollar industry; it is labour which 'ensure[s] that Grandma never has to see images like the one Babayan just nuked' (2014). Leaving this image available for users to view, it is argued, would not only turn uses away from the site, it would create a user experience that is disruptive, unpleasant, and traumatic. As such, someone—7,000 miles away, in a former colony, in a formerly used school—is employed to erase it. But what is the work that this erasing does? Can it really be said that this outsourced labour successfully mitigates the trauma produced through the witnessing of an image such as this? In what follows, I begin to answer these questions along two lines of analysis. First, I turn to the scholarship on witnessing to complicate the very demand that 'traumatic' images go unseen. Second, I argue that due to the very nature of the labour of content moderation, the assumption that erasing this image is successful in mitigating the alleged harm of the image cannot be sustained. It is precisely because this harm cannot be contained and is reproduced through the very labour of erasing it, I argue, that this labour is relegated to invisibilized workers living in the margins.

To make sense of this argument, it is necessary to turn to scholarship on the ethics of witnessing. In the wake of images and testimonies of death and torture of those who were forcibly sent to concentration camps during the Holocaust, a longstanding area of philosophical questioning has centred around what it means to be a witness to horror, violence, and suffering (Agamben 1999; Felman and Laub 1992; Sontag 2003). The scholarship on the ethics of being a witness to trauma has since expanded to both challenge the centrality of the Holocaust as the exceptional European urtext of unimaginable violence (Hirsch and Spitzer 2009; Levi 2007; Levy and Sznaider 2006;

Rothberg 2009), and to interrogate the act of witnessing in relation to other violences of many kinds: the trauma of colonial dispossession (Craps and Buelens 2008; Hutcheon 2003), the September 11<sup>th</sup> terror attacks (Zelizer 2002), images of lynching (Polchin 2007), domestic violence (Henke 1998; Herman 1992; Kilby 2007), and the torture at Abu Ghraib (Feldman 2005; Puar 2004; Mirzoeff 2006). Within this scholarship, one of the central ethical questions of bearing witness is the shift from the individualized act of seeing an image, or listening to a testimony, and what Barbie Zelizer describes as the 'adoption of a public stance by which they [individuals] become part of a collective working through trauma together' (2002, 52).

In a foundational sense, content moderators, and the companies which employ them and which set the guidelines for their labor, are already knowingly aware of their work's ethical relationship to the act of witnessing. Buni and Chemaly use the language of bearing witness, for example, to write about the swift policy shifts that were undertaken at YouTube in the midst of the 2009 protests in Iran against the presidential victory of Mahmoud Ahmadinejad:

On June 20th [2009], the [YouTube SQUAD] team was confronted with a video depicting the death of a young woman named Neda Agha-Soltan. The 26-year-old had been struck by a single bullet to the chest during demonstrations against progovernment forces and a shaky cell-phone video captured her horrific last moments: in it, blood pours from her eyes, pooling beneath her. Within hours of the video's upload, it became a focal point for Mora-Blanco [a former YouTube employee] and her team. As she recalls, the guidelines they'd developed offered no clear directives regarding what constituted newsworthiness or what, in essence, constituted ethical journalism involving graphic content and the depiction of death. But she knew the video had political significance and was aware that their decision would contribute to its relevance. Mora-Blanco and her colleagues ultimately agreed to keep the video up. (2016)

Similar stories are relayed by various reporters and content moderators, about whether or not moderators should allow or delete images and footage of the Arab Spring, police killings of black people, people in blackface, beheadings, as well as, according to a leaked Facebook moderation document, posts by Kurdish users who are critical of the Turkish government (Chen 2017). As with the decision about the footage of Agha-Soltan's death, for the most part, violent imagery that companies allow to be an exception to their policy's bans are justified based on the content's 'newsworthiness.' In a telling passage that links this very journalistic politics of witnessing with a neoliberal panic surrounding the democratization of content in the digital era, Buni and Chemaly relay YouTube's logic for allowing the footage:

It [the video] was fuelling important conversations about free speech and human rights on a global scale and was quickly turning into a viral symbol of the movement. It had tremendous political power. *They* had tremendous political power. And the clip was already available elsewhere, driving massive traffic to competing platforms. (2016)

In commercial content moderation, decisions about what to show, and what exceptions to make, are thus not just tied to the ethics of the image, they are also explicitly linked to brand management and neoliberal competition. In this context, Agha-Soltan's death literally becomes a product which YouTube cannot afford not to sell. vii

Nonetheless, the questions that are asked here about witnessing are important ones, particularly as the immediacy of the digital image, and the global reach of social media, intensifies what Ariella Azoulay calls the photograph's potential for political 'encounter' (2008). Azoulay, writing about the increasingly 'widespread use of cameras by people around the world', argues that the growing multitude of images has 'created a new form of encounter, an encounter between people who take, watch, and show other people's photographs, with or without their consent, thus opening new possibilities of

political action and forming new conditions for its visibility' (2008, 24). It is precisely this twinned effect of the image—allowing the bearing of witness to trauma, and opening up new forms of alliance—which allegedly underscores the complex labor of human moderation. 'Nailing down the ineffable question of why one piece of content is acceptable but a slight variation breaks policy', Buni and Chemaly write, 'remains the holy grail of moderation' (2016). It is the difficulty in settling this precise yet murky distinction which requires human, rather than algorithmic, labor.

What is consistent across these accounts of companies negotiating the ethics of the image, however, is that they only pertain to a particular form of political violence. No debate appears to be had, that is, about what the ethics of witnessing sexual violence or sexual expression are. In a post titled 'Facebook's Community Standards: How and Where We Draw the Line' (2017), Monika Bickert, the Head of Global Policy Management at Facebook, explains the complexities of content moderation. In her explanation, however, the practices for moderating violent footage of a chemical weapons attack are given context and justification, while the logics behind regulating sexuality are mostly taken for granted. Bickert writes:

Last month, people shared several horrific videos on Facebook of Syrian children in the aftermath of a chemical weapons attack. The videos, which also appeared elsewhere on the internet, showed the children shaking, struggling to breathe and eventually dying. The images were deeply shocking – so much so that we placed a warning screen in front of them. But the images also prompted international outrage and renewed attention on the plight of Syrians. Reviewing online material on a global scale is challenging and essential. ... The cases we review aren't the easy ones: they are often in a grey area where people disagree. Art and pornography aren't always easily distinguished, but we've found that digitally generated images of nudity are more likely to be pornographic than handmade ones, so our policy reflects that. (2017)

Echoing the discussion of the merits of hosting the video of Agha-Soltan's death above, here, Bickert acknowledges both the journalistic merit of hosting violent footage of warfare, and the necessity of remaining competitive. Complex thought seems to be going into the debate about whether or not the potential trauma that this footage might evoke is justified by the urgency of the circumstances of its production and dissemination. In contrast, the realm of the pornographic is given no consideration as such. Knowingly pornographic, and knowingly inappropriate, the material which blurs the line between art and pornography is deemed meritless. Here, distinguishing between art and pornography is understood to be a complex act, and yet, the implicit implication of assessing something as pornography is that it should therefore necessarily be removed. There are seemingly no instances in which sexual content is understood as politically useful to show. Unlike (political) violence, pornography (however defined) is not understood to have social or political value. Viii

While this may, of course, be understood as simply a brand-management question, or a legal one—with anti-obscenity laws prohibiting the displaying, hosting, and access to certain sexual images across various websites and apps—there are two important things to keep in mind here. The first is that, as feminist scholars have argued for decades (Brownmiller 1975; Crenshaw 1993; Kappeler 1986; Kilby 2007; Woodiwiss 2014), there are important, albeit difficult, questions of testimony, witnessing, and accountability, which accompany narratives and images of sexual abuse and survivorship. For Lynn Higgins and Brenda Silver, editors of the collection *Rape and Representation* (1991), the questions of witnessing and testimony are indeed central to understanding what is currently referred to as 'rape culture'. They argue that testimony, in particular, is fundamental to the ways in which rape and sexual violence 'have been so ingrained and so rationalized through their representations' (1991, 2),

because women's own testimonies about surviving rape are so often discredited. For Higgins and Silver, the fact that 'representations of rape after the event are almost always framed by a masculine perspective premised on men's fantasies about female sexuality and their fears of accusation, as well as their codified access to and possession of women's bodies' (2), necessitates seeking out new 'rhetorical strategies whereby rape gets represented in spite of (or through) its suppression' (4). They thus argue for a critical commitment to being a witness to rape testimony, and to taking it seriously as evidence of a physical, sexual violation that is a product of the 'obsessive inscription—and [the] obsessive erasure—of sexual violence against women and those placed by society in the position of "woman" (2). Being a critically engaged and generous witness to this testimony, in other words, opens up possibilities for empathy, trust, and accountability.

While Higgins and Silver are thus interested in thinking through the potentials of 'rereading rape' at the textual level, Tanya Horeck's work on representations of rape, real and fictionalized, situates the politics of witnessing rape and sexual violence in relation to the image. Horeck (2004) specifically addresses the 1983 gang rape of a twenty-one year-old woman at a bar called Big Dan's, which occurred as the patrons of the bar cheered, laughed, and goaded it on. The ensuing rape trials—at which the defense attorneys used the survivor's sexual history to blame her for her own rape—were among the first ever court cases to be nationally broadcast live on television. Both the rape itself, and the trials, additionally found a site of re-presentation, as they became the basis for a 1988 feature film, *The Accused*. Despite what might be understood as an excess of representation, one of Horeck's central arguments about the Big Dan's rape case, is that the initial rape itself was premised on a failure of witnessing; the bar patrons relayed their proximity to the rape as mere 'spectators', not as 'witnesses'. None

of them called for help or intervened. As such, the broadcasting, and re-presenting of the rape, both on live television and in fictionalized filmic form, signalled a particular paradox: 'a case that received extensive publicity because it exposed a communal failure to witness a woman's rape, inaugurates a form of representation premised on the idea that communal looking serves the ends of civic justice' (2004, 85). Horeck, drawing upon Ruby Rich (1983), asks: 'how can looking cure the damage caused by looking?' (97). This case illuminates, Horeck argues, that there is a fundamental 'crisis' at the heart of witnessing. While witnessing is premised on, and perhaps allows for, communal empathy and accountability, it often requires the very same act—that of looking—which was foundational to the initial violence itself. As such, witnessing is not a straightforward act of accountability or ethics. Rather, it is a deeply ambivalent act which can simultaneously engender politically mobilizing active empathy, and engender violent participation.

Linking these two accounts back to content moderation, it is clear that the notion that flagged pornographic imagery is already knowingly unethical, and thus does not require a conversation about bearing witness to it, elides both the ethical imperative that witnessing and testimony call forth, and the ambivalences around looking which are central to the image. I need to be careful here in how I articulate this point, as my intention is not to flatten out all pornography, or indeed all sexual imagery (for flagged digital content is capacious in its breadth and intensity) as somehow equally and unequivocally violent. Nor am I suggesting that Facebook, for example, is necessarily the most appropriate place to hold such a debate. However, rather than attempt to resolve these difficulties here, I want instead to raise the second element which is central to the ways in which content moderation addresses these very questions. It is important to remember, that is, that while these images are removed from view for *users* 

of these websites and apps—and thus these questions of testimony, witnessing, and looking are evaded, at least somewhat, for 'us'—there are still people around the globe, like Babayan, who view, review, and re-view, these images on a daily basis. These very questions of looking and witnessing, in other words, are not resolved by the offshoring of content moderation to various outsourcing companies in the Philippines. The act of looking is in fact absolutely central to their daily labor. As such, rather than argue for a straightforward resolution to the 'crisis' in witnessing, I am arguing that the particular post-colonial relation that is facilitated by content moderation instigates a form of labor that displaces the very potential for witnessing to an elsewhere that sits 'beyond' the nation but firmly within the grasp of the metroimperial. The relegation of content moderation to invisibilized precarious labourers in the Global South is thus, I am arguing, less about the political economy of content moderation, as social media sites might claim, and is instead precisely about disavowing and displacing acts of witness and their inherent potential for instigating empathy and accountability onto a post-colonial other.

#### **Conclusion: Witnessing Elsewhere**

Returning to Zelizer's argument that one of the preconditions for witnessing (rather than spectating) is the formation of a collective who works through trauma together, it is important to consider that because the labor of moderating takes place in an invisibilized 'elsewhere,' no collective, beyond these laborers, can form. Even for these laborers, collectivity seems precluded:

Without visible consequences here and largely unseen, companies dump child abuse and pornography, crush porn, animal cruelty, acts of terror, and executions — images so extreme those paid to view them won't even describe them in words to their loved ones — onto people desperate for work. And there they sit in

crowded rooms at call centers, or alone, working off-site behind their screens and facing cyber-reality, as it is being created. (Buni and Chemaly 2016)

There is no way to know exactly what the consequences of the displacement of witnessing into siloed, marginalized communities 'elsewhere' might be for the pervasiveness of sexual violence in the US, or for the possibilities and limitations of transpacific alliances premised on survivorship, accountability, and harm-prevention. Nor, unfortunately, is definitively forging these links possible within the realm of this article. Rather, I conclude with two suggestions for future research. First, one of the implications of my argument for centring the labor of content moderation within porn studies is the recognition that in the digital age, coming down on the prohibition side of the sex wars debate does not end the ambivalence of looking relations. The flagging, filtering, and prohibiting of digital sexual content may dramatically limit the reach of that content 'here', but the work of cleaning up the internet (a global care chain of a different kind) simply displaces where the effects of these images are felt, not that they are. Second, as I have been arguing, the implications of studying pornography under what I have been calling the 'digital life of coloniality' require that we think anew about the multiple directions and effects of power, subjectivity, and labor at play in regulating the proliferation of sexualities online.

.

<sup>&</sup>lt;sup>i</sup> According to Monika Bickert, Facebook's head of global product policy, Facebook users flagged more than one million items of content for review every day in 2016 (Buni and Chemaly 2016).

ii PhotoDNA is a technology developed by Microsoft that uses a technique called 'robust hashing' which creates a digital fingerprint for an image. This then matches it to other copies of that same image, even if the image has been altered. It was developed to track the spread of child pornography across the internet.

iii Content moderation is more of a fragmented process than I am able to account for here. It includes workers located within in-house departments at various technology companies, outsourcing firms, call centres, and micro-labor sites, as well as 'untold numbers of algorithmic and automated products' (Buni and Chemaly 2016).

- <sup>iv</sup> While content moderation is thus a form of outsourced labor, it is not exactly the same job as working in a call center, the more traditional understanding of this type of work. Despite their differences, however, both of these industries can be understood together though an analysis of their shared political economy, and their emergence in certain geopolitical locations, from specific colonial relations.
- <sup>v</sup> For other analyses of the American colonial administration of the Philippines, see, among others: Go (2008); Rafael (2000); Westling (2011).
- vi Eventually, Buni and Chemaly write, 'the "stew pot guy" began uploading more explicit content that clearly violated Pinterest's terms and [the] team removed his account' (2016).
- vii For an analysis of the circulation of the footage of Agha-Soltan's death, one which questions why this footage of death, rather than, say, the footage of Oscar Grant's murder, functioned as a global catalyst for action and empathy, see: Malkowski (2017).
- between pornography, sexual imagery, sexual expression, and sexual content. While I agree with Bickert here that careful analysis and contextualization is necessary for such an act, I hesitate to make such a distinction within this piece precisely because of the ways in which, as within this quote from Bickert, the lines between art and pornography are blurred by social media companies. Rather than attempt to resolve slippages in my own writing between these different (yet overlapping) formulations, I am seeking to emphasise their slipperiness. For it is precisely the ways in which they are produced as slippery signifiers (mundane sexual expression for one becomes pornography for another) that they are able to take on so much work in terms of affect, accountability, and politicization.
- ix While one could thus argue that content moderators' reviewing and re-viewing of images of sexual abuse, particularly child sexual abuse, might engender re-traumatization—particularly given that one of the central arguments made against child pornography is that the child's knowledge of the image's circulation and viewing by others is in and of itself a form of trauma (Smolen 2013; Oswell 2006)—this is not my intent here.

#### References:

- Agamben, Giorgio. 1999. Remnants of Auschwitz: The Witness and the Archive. New York: Zone Books.
- Alexander, Jacqui M. 1994. "Not Just (Any) Body Can Be a Citizen: The Politics of Law, Sexuality and Postcoloniality in Trinidad and Tobago and the Bahamas" *Feminist Review* 48: 5-23.
- Azoulay, Ariella. 2008. The Civil Contract of Photography. New York: Zone Books.
- Barnett, Emma and Hollingshead, Iain. 2012. "The dark side of Facebook" *Telegraph*, March 2.
- Bhabha, Homi. 1991. The Location of Culture. New York: Routledge.
- Bickert, Monica. 2017. "Facebook's Community Standards: How and Where We Draw the Line" *Facebook Newsroom*, May 23.
- Brownmiller, Susan. 1975. *Against Our Will: Men, Women and Rape*. London: Penguin.
- Buni, Catherine and Chamaly, Soraya. 2016. "The secret rules of the internet" *The Verge*, April 13.
- Chen, Adrian. 2014. "The Laborers Who Keep Dick Pics and Beheadings Out of Your Facebook Feed" *Wired*, October 23.
- Chen, Adrian. 2017. "The Human Toll of Protecting the Internet From the Worst of Humanity" *The New Yorker*, January 28.
- Chokshi, Nirtaj. 2017. "YouTube Filtering Draws Ire of Gay and Transgender Creators" *New York Times*, March 20.
- Craps, Stef and Buelens, Gert, eds. 2008. "Postcolonial Trauma Novels" Special Issue *Studies in the Novel* 40 (1&2).
- Crenshaw, Kimberlé. 1993. "Whose Story Is It Anyway? Feminist and Antiracist Appropriations of Anita Hill" in Toni Morrison, ed. *Race-ing Justice, Engendering Power. Essays on Anita Hill, Clarence Thomas, and the Construction of Social Reality.* London: Chatto, 402-440.
- David, Emmanuel. 2016. "Outsourced Heroes and Queer Incorporations: Labor Brokerage and the Politics of Inclusion in the Philippine Call Center Industry" *GLQ* 22 (3): 381-408.
- Fanon, Frantz. 1967. *Black Skin, White Masks*. Translated by Charles Lam Markmann. London: Grove Press.
- Feldman, Allen. 2005. "On the Actuarial Gaze" Cultural Studies 19 (2): 203-226.

- Felman, Shoshana and Laub, Dori. 1992. *Testimony: Crises of Witnessing in Literature, and History*. New York: Routledge.
- Go, Julian. 2008. American Empire and the Politics of Meaning: Elite Political Cultures in the Philippines and Puerto Rico during U.S. Colonialism. Durham: Duke University Press.
- Henke, Suzette. 1998. Shattered Subjects: Trauma and Testimony in Women's Life-Writing. Basingstoke: Macmillan.
- Herman, J. Lewis. 1992. *Trauma and Recovery: From Domestic Abuse to Political Terror*. London: Pandora.
- Higgins, Lynn and Silver, Brenda. 1991. "Introduction: Rereading Rape" in *Rape and Representation*. New York: Columbia University Press.
- Hirsch, Marianne and Spitzer, Leo. 2009. "The witness in the archive: Holocaust Studies/Memory Studies" *Memory Studies* 2 (2): 151-170.
- Horeck, Tanya. 2004. *Public Rape: Representing Violation in Fiction and Film.*London: Routledge.
- Hutcheon, Linda. 2003. "Postcolonial Witnessing and Beyond: Rethinking Literary History Today" *Neohelicon* 30 (1): 13-30.
- Isaac, Punzalan Isaac. 2017 "Offshore Identities: Ruptures in the 300-Second Average Handling Time" in Brian Russell Roberts and Michelle Ann Stephens, eds. *Archipelagic American Studies*. Durham: Duke University Press.
- Kappeler, Susanne. 1986. The Pornography of Representation. Cambridge: Polity Press.
- Kilby, Jane. 2007. *Violence and the Cultural Politics of Trauma*. Edinburgh: Edinburgh University Press.
- Levi, Neil. 2007. "No Sensible Comparison?' The Place of the Holocaust in Australia's History Wars" *History and Memory* 19 (1): 124-156.
- Levine, Philippa. 2000. "Orientalist Sociology and the Creation of Colonial Sexualities" *Feminist Review* 65: 5-21.
- Levy, Daniel and Sznaider, Natan. 2006. *Holocaust and Memory in the Golden Age*. Philadelphia: Temple University Press.
- Malkowski, Jennifer. 2017. *Dying in Full Detail: Mortality and Digital Documentary*. Durham: Duke University Press.
- McClintock, Anne. 1995. *Imperial Leather: Race, Gender and Sexuality in the Colonial Context*. New York: Routledge.

- Mendoza, Victor Román. 2015. *Metroimperial Intimacies: Fantasy, Racial-Sexual Governance, and the Philippines in U.S. Imperialism, 1899-1913*. Durham: Duke University Press.
- Mirchandani, Kiran. 2012. *Phone Clones: Authenticity Work in the Transnational Service Economy*. London: Cornell University Press.
- Mirzoeff, Nicholas. 2006. "Invisible Empire: Visual Culture, Embodied Spectacle, and Abu Ghraib" *Radical History Review* 95 (spring): 21-44.
- Mulholland, Monique. 2016. "The pathological native' versus 'the good white girl': an analysis of race and colonialism in two Australian porn panics" *Porn Studies* 3 (1): 34-49.
- Oswell, David. 2006. "When images matter: Internet child pornography, forms of observation and an ethics of the virtual" *Information, Communication and Society* 9 (2): 244-265.
- Polchin, James. 2007. "Not looking at Lynching Photographs" in Frances Guerin and Roger Hallas, eds. *The Image and the Witness: Trauma, Memory and Visual Culture*. London: Wallflower Press.
- Povinelli, Elizabeth. 1997. "Sex Acts and Sovereignty: Race and Sexuality in the Construction of the Australian Nation" in R. N. Lancaster and M. di Leonardo, eds. *The Gender/Sexuality Reader: Culture, History, Political Economy*. New York: Routledge.
- Puar, Jasbir. 2004. "Abu Ghraib: Arguing against Exceptionalism" *Feminist Studies* 30 (2): 522-534.
- Puri, Jyoti. 2014. "States' Sexualities: Theorizing Sexuality, Gender and Governance" in Mary Evans et al., eds. *Handbook of Feminist Theory*. London: Sage. pp. 343-358.
- Rafael, Vicente L. 2000. White Love and Other Events in Filipino History. Durham: Duke University Press.
- Rich, Ruby. 1983. "Anti-Porn: Soft Issue, Hard World" Feminist Review 13: 56-67.
- Roberts, Sarah T. 2016. "Digital Refuse: Canadian Garbage, Commercial Content Moderation and the Global Circulation of Social Media's Waste" *Media Studies Publications*. Paper 14.
- Rotherberg, Michael. 2009. *Multidirectional Memory: Remembering the Holocaust in the Age of Decolonization*. Stanford: Stanford University Press.

- Rubin, Gayle. 1984 "Thinking Sex: Notes for a Radical Theory of the Politics of Sexuality" in Henry Abelove and David Halperin, eds. (1993) *The Lesbian and Gay Studies Reader*. London: Routledge.
- Smolen, Molly. 2013. "Redressing Transgression: In Defense of the Federal Sentencing Guidelines for Child Pornography Possession" *Berkeley Journal of Criminal Law* 18 (1): 36-83.
- Sontag, Susan. 2003. *Regarding the Pain of Others*. New York: Farrar, Straus and Giroux.
- Spivak, Gayatri Chakravorty. 1988. "Can the Subaltern Speak?" in Cary Nelson and Lawrence Grossberg, eds. *Marxism and the Interpretation of Culture*. London: Macmillan.
- Stoler, Ann Laura. 1995. *Race and the Education of Desire: Foucault's* History of Sexuality *and the Colonial Order of Things*. Durham: Duke University Press.
- Stone, Brad. 2010. "Policing the Web's Lurid Precincts" New York Times, July 18.
- Westling, Meg. 2011. *Empire's Proxy: American Literature and U.S. Imperialism in the Philippines*. New York: New York University Press.
- Woodiwiss, Jo. 2014. "Beyond a single story: The importance of separating 'harm' from 'wrongfulness' and 'sexual innocence' from 'childhood' in contemporary narratives of childhood sexual abuse" *Sexualities* 17 (1/2): 139-158.
- Woollacott, Angela. 1999. "White colonialism and sexual modernity: Australian women in the early twentieth century metropolis" in Antoinette Burton, ed. *Gender, Sexuality and Colonial Modernities*. London: Routledge.
- Zelizer, Barbie. 2002. "Photography, journalism, and trauma" in Barbie Zelizer and Stuart Allan, eds. *Journalism After September 11*. London: Routledge.