Book Review

new media & society
1–3

DOI: 10.1177/1461444819878844
journals.sagepub.com/home/nms
SAGE

Sarah T. Roberts, *Behind the Screen: Content Moderation in the Shadows of Social Media*. Yale University Press: New Haven, CT, 2019; 280 pp.; ISBN: 9780300235883, $30.00 (hbk)

Reviewed by: Ysabel Gerrard, *University of Sheffield, UK*

'Out of the shadows, out from behind the screen, and into the light'. (p. 222)

Sarah T. Roberts' *Behind the Screen: Content Moderation in the Shadows of Social Media* offers a groundbreaking exploration of commercial content moderation (CCM): a term the author coined to denote the essential yet taken-for-granted human labour of 'screen[ing] content uploaded to the Internet's social media sites on behalf of the firms that solicit user participation' (p. 1). Although Internet communities have long been overseen by volunteer moderators, Roberts explains that content moderation is now done *at scale*: it is a (lowly) paid role at some of the world's largest commercial entities. Human content moderation usually takes place after a piece of content (an image, video, comment, etc.) has been uploaded to social media, when a user 'flags' something because they think it breaks the rules. Until fairly recently, and thanks in large part to Roberts' research, few social media users knew their complaints were directed to the humans working in what the author calls the *shadows of social media*. Published in 2019 by Yale University Press, *Behind the Screen* is the unique culmination of almost a decade's worth of Roberts' groundbreaking research (see also Roberts, 2014, 2016, 2017a, 2017b, 2018a, 2018b). Drawing on interview data from field sites spanning the Philippines, rural Iowa, Silicon Valley, India, Canada and Mexico, *Behind the Screen* offers a stunning and at times harrowing insight into the working lives of current and former CCMs: the distressing content they endure, their unstable working conditions and, perhaps most significantly, the declining state of their mental health.

In Chapter 1, Roberts traces the rise of public knowledge about CCM, including the beginnings of her own research. Roberts shows how she has helped to bring CCM into the public agenda, raising awareness of 'the fraught and difficult nature of such front-line online screening work' (p. 3). Next, Chapter 2 offers a theoretically driven and useful description of CCM, the contexts in which it takes place and its position within historical and contemporary discussions of labour relations. Roberts explains that CCM is 'fractured organizationally and geographically' (p. 39). For example, *in-house* moderators

operate on-site at the company they work for, *boutique* moderators are recruited by firms specialising in brand management services, *call-centre* moderators work for large-scale organisations offering 24/7 content moderation services to large social media companies and *microlabor platform* moderators are geographically dispersed workers who engage in 'digital piecework': content moderation paid on a per-task basis (pp. 39–48). The fragmentation of the CCM workforce raises a number of concerns for Roberts, which she discusses at length in Chapters 3, 4 and 5.

Chapter 3 is the first of three empirically grounded chapters to tell the powerful and largely unheard stories of content moderators working in-house at social media companies' Silicon Valley headquarters. Although companies tend not to offer permanent CCM roles, the prestige of working in the Valley is enough to draw workers in. The author takes this opportunity to remind readers that content moderation guidelines are developed 'in the specific and rarefied sociocultural context of educated, economically elite, politically libertarian, and racially monochromatic Silicon Valley, USA' (pp. 93–94): crucial framing for the rest of the book. Roberts' informants told her that they grapple with biases within companies' strict content moderation guidelines. For example, one CCM worker, Max, was distressed to find that 'blackface is not technically considered hate speech by default' (p. 94) at MegaTech: his pseudonymously named Silicon Valley employer.

The topic of corporate and cultural conformity continues into Chapter 4, where Roberts shares the stories of boutique and microlabor platform workers. These forms of moderation differ from the in-house moderation discussed in Chapter 3; for example, boutique and microlabor platform workers are almost exclusively stationed at home, often work on news sites and tend not to witness graphic imagery. But all the moderators discussed so far share difficulties in jumping between different cultural contexts and client 'voices' (p. 142). As Rick explained, 'Some of the news sites might be pro-NRA, some might be pro-gun control, so you need to put aside your personal beliefs and moderate, not "I think that comment is appropriate or not appropriate"' (p. 145).

Conforming to a site's 'voice' is a demand that seems to be placed on all CCMs, including call-centre workers based in the Philippines. These workers engage in outsourced CCM labour from some of the largest social media companies in the world and have 'just seconds' to 'review, edit, delete, [and] resolve' cases, or 'tickets' as they are known in the industry (p. 173). Similar to in-house moderation, call-centre CCM work is heavily metricised and workers can get through thousands of tickets per shift. Chapter 5 offers vital historical context for digital labour in the Philippines: a country that recently overtook India as the call-centre capital of the world, despite having only one-tenth of its population (p. 181). These workers are required to undertake moderation outside of their 'spoken language of choice and everyday cultural context': an 'extremely difficult' task that 'presents novel and distinct challenges' (p. 194).

Roberts' empirically based chapters (3, 4 and 5) helpfully discuss the differences between each category of CCM work, but what unites all workers is the prevalence of mental health issues. This is one of the most significant revelations in *Behind the Screen* – and newsworthy, hence the press Roberts' research has received. Although Roberts never directly asked her informants to talk about their worst experiences – an admirably ethical decision – she notes that examples inevitably surfaced during her interviews: child and animal abuse, hate speech, graphic sexual content, war zone footage, to name

only some of the worst things the participatory Web has to offer. By placing CCM workers on short-term contracts and denying them access to the healthcare benefits enjoyed by permanent employees, large social media companies are enacting 'a series of distancing moves designed to create a plausible deniability to limit their responsibility for workplace harm, particularly when such harm may take time to show up' (p. 127). Roberts notes an absence of studies on the long-term effects of CCM work, but *Behind the Screen*'s powerful role in bringing content moderation 'into the light' (p. 222) will hopefully prompt some of this urgently needed research.

Roberts' exemplary research has helped to shape and has been shaped by a growing body of scholarly work on content moderation. Accessibly written, *Behind the Screen* is essential reading for anyone – students, academics, journalists, policy makers and practitioners – with a desire to make better sense of the relationship between social media and society. But where do we go from here, now the lid has been lifted on social media's shadowy workforce? In the book's concluding chapter, Roberts explains why social media companies will likely always need to rely on human content moderators, and with this in mind, she leaves readers to consider a series of questions about the future of CCM work.

Few academic books play out so publicly and so powerfully like *Behind the Screen*. The book's underlying research has helped to inform academic studies, corporate policies, investigative journalism, documentary films, lawsuits filed by CCM workers, and activism and civil society interventions. *Behind the Screen* has helped to bring CCM 'out of the shadows, out from behind the screen, and into the light' (p. 222), where I sincerely hope it remains.

ORCID iD

Ysabel Gerrard https://orcid.org/0000-0003-1298-9365

References

Roberts ST (2014) *Behind the screen: the hidden labor of commercial content moderation*. PhD Dissertation, University of Illinois at Urbana–Champaign, Urbana, IL.

Roberts ST (2016) Commercial content moderation: digital laborers' dirty work. In: Noble SU and Tynes B (eds) *The Intersectional Internet: Race, Sex, Class and Culture Online*. New York: Peter Lang, pp. 147–159.

Roberts ST (2017a) Aggregating the unseen. In: Byström A and Soda M (eds) *Pics or It Didn't Happen*. Munich: Prestel, pp. 17–21.

Roberts ST (2017b) Social media's silent filter. *The Atlantic*, 8 March. Available at: https://www.theatlantic.com/technology/archive/2017/03/commercial-content-moderation/518796/

Roberts ST (2018a) Digital detritus: 'error' and the logic of opacity in social media content moderation. *First Monday* 23(3). Available at: http://firstmonday.org/ojs/index.php/fm/article/view/8283/6649

Roberts ST (2018b) Meet the people who scar themselves to clean up our social media networks. *Maclean's*, 15 June. Available at: https://www.macleans.ca/opinion/meet-the-people-who-scar-themselves-to-clean-up-our-social-media-networks/

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