

## Tarleton Gillespie

*Custodians of the Internet. Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*, New Haven, Yale University Press, 2018, pp. 296

### Corinna Canali Royal Academy of Art, The Hague

Not everything posted on social media makes it to the public: some bits of content disappear along the way and who makes them disappear are the platforms themselves. Several times Facebook (but not only), has been under the spotlight for taking down or allowing specific content posted by their users. The latest episode saw Mark Zuckerberg's main platform removing a picture of the Venus of Willendorf, causing the outrage of the Naturhistorische Museum (NHM) of Vienna, where the Venus is physically located. After the episode was brought under public attention, Facebook restored the post and apologized (Dawson 2018).

Ever since they emerged from the fabric of the web 2.0, social media companies have always presented themselves as mere conduits of content, pushing afar every responsibility on what ended up on the spaces they provided. But such a portrait is contradicted by the moderating systems they apply on users.

After the breach opened by the unprecedented studies of Sarah T. Roberts, assistant professor in the UCLA Department of Information Studies, revealing the logics moving complex systems of social media moderation, Tarleton Gillespie in his book *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions that Shape Social Media*, published by Yale University Press in 2018, engages in an extensive analysis of the platforms' thorough "governance mechanisms", as James Grimmelman (2015) defined them. Gillespie, principal researcher at Microsoft Research New England and veteran in a research field that links media studies to technology and science studies (see Gillespie et al. 2014), tells about moderating systems in a way social media do not, i.e. with transparency. Kept confusing and mostly untold, such systems are enacted with users not even noticing them and totally incapable of accessing them. Through the words of Gillespie, the encounter with moderating mechanisms is finally made accessible and clear, brought to its very essence.

The book follows a rather linear path. Opening with the infamous removal of Nick Ut's photography known as "Napalm Girl" from several accounts of Norwegian citizens and politicians, Gillespie, wisely unfolds throughout his investigation, a detailed explanation of what CCM is, who and what are its main actors and why it is a very problematic mechanism. Revealed across eight chapters, what the author presents is a wide spec-

trum of not only technical and legal features, but most of all, of the reasons why the entire system of online moderation as it is now requires a deep change.

It is emblematic that such book begins recounting a debated removal. Content moderation, technically defined Commercial Content Moderation (CCM), is an untold and invisible system, moving platforms that independently define what is to be made public and what is not based on doubtful logics (Roberts 2016). Hidden behind unaccessible mechanisms, moderation is hard to spot. It reveals itself in the act of taking down user-generated content (UCG), exposing the power enacted by platforms providers. Meaning that, unless users are posting forbidden material, they will hardly even notice CCM at work. But this does not mean that they are not subject to it.

Content moderation is the infrastructure supporting and moulding social media spaces and, as for all infrastructures, its very nature is invisibility. It shows itself “upon breakdown”, meaning that it becomes visible when it breaks (Star 1999). As Gillespie explains, despite being a fundamental feature infused in almost the entirety of Western-based social media platforms, it is still unknown to the majority of social media users. As said, using platforms as commanded by their regulating norms, users never encounter moderation at all. And those who do, are the ones that, acting against its commands, get cut out, silenced or, borrowing social media logos, banned.

“Custodians of the Internet” not only engages in an in-depth explanation of the system itself: it points out the various issues it generates. Every infrastructure unfolds specific narrations (Star 1999), and so does moderation. The core action of CCM is to allow or forbid pieces of content. And the judgment shapes what the platform is and what it is not. The list of prohibitions, entirely decided by the private companies owning such platform, changes at will, and empowers tech companies on deciding to set rules along the way.

In the first and second chapters, Gillespie presents evidence of both the inherited non-neutrality of social media platforms and the wider struggles to regulate the Internet that, already in the first 1990s, saw lawmakers and Internet *connoisseurs* facing the dilemma of setting boundaries to online activity without jeopardising users’ freedom of expression. The norms regulating UCG are key elements in CCM systems and the third chapter focuses on their role. Presenting as case-studies analyses of snippets taken from guidelines regulating various platforms, Gillespie demonstrates the reasons moving them, how they are constructed and what are the similarities between different platform providers moved by similar motives.

And if so far the problematic aspects were only starting to surface, from the fourth chapter on Gillespie engages in a deeper explanation of the issues surrounding and emerging from attempts at moderating global

scale content. Starting from how both traditional and social media decide to moderate content, the fourth chapter presents forms of moderation and their interconnected issues. Shifting to the fundamental human labour shaping CCM, the fifth chapter analyses how moderators are trained and the working conditions they face. Furthermore, Gillespie questions the fairness of a global scale moderation managed by a small specific group of individuals mainly based in the Silicon Valley area.

Fairness concerns come back when, in the sixth chapter, the author moves from an analysis tackled from the perspective of platforms and their managers to the users' one. Presenting both (in)famous and less known evidence of how social media moderation can become a tool for discrimination, in this section the author wisely confronts female objectification and gender discrimination enforced through moderating systems. Using it to prove the tensions that arise when users confront platforms policies they can hardly appeal. Tensions that expose the total discretion of the platform on what/why/how to regulate. Gillespie uncovers the subjectiveness moving policies using as an example how Facebook, only after a long set of public contestations, tweaked its female nudity rules shifting from an absolute ban of depictions of female breasts to allowing them in restricted specific cases (e.g. breastfeeding).

The final two chapters begin presenting possible solutions to the problem, with the seventh chapter focussing on the question whether it is better to remove content or rather filter and hide it, giving actual examples of both approaches. Note that the book uses Tumblr policy as an example of alternative moderating systems, where adult content was filtered instead of removed altogether. But, by the time it was published, the platform's permissive policy was replaced by a stricter and more conformed one a few months after the enactment of the "Stop Enabling Sex Traffickers Act" (SESTA) and "Allow States and Victims to Fight Online Sex Trafficking Act" (FOSTA) regulation (see Romano 2018).

The last chapter explicitly looks at the nature of tech companies providing social media spaces, questioning the liberties that Western societies have allowed them and including the users' responsibility towards such allowance. The conclusory part of Gillespie's analysis presents detailed possibilities to improve social media starting from how, what and why they should moderate challenging how social media companies have positioned themselves both on- and offline.

Gillespie, through his attentive analysis, warns us of the dangers of such empowerment when he tells that our public culture is, in important ways, shaped and designed by the platforms we access. When the encounter user-technology happens, is the latter that pulls the strings. Users are constantly re-configured and educated, pushed into tight boundaries designed to preserve the perfected version of reality infused in the machine (Woolgar 1991). Whoever fails at conforming, is banned and silenced. And if users are the raw material to be (re)configured, norms are

the mould. Such massively frequented spaces are the tools used to “normalise” humans, through prototypical structures affecting all the non-conformed ones (Browne 2015).

*Custodians of the Internet* explains why we, active or inactive users, should question the promised impartiality of such powerful sites. Moderation, in this investigation, turns out to be not only an optional feature relegated to a few strict social media platforms. It is the central value proposition for all of them. Their very commodity subtly based on and used to shape users’ engagement and participation, attracting or repelling investors accordingly.

Moderation is the core feature of online social networks and we should demand for transparency in the system. Designing systems to moderate user-generated material is social media companies’ main occupation—only at Facebook, Inc. it involves some 30.000 individuals, half of which are moderators. Furthermore, users should engage in understanding and questioning such systems as they are actively part of it. As *Custodians of the Internet* explains, the “custodians” are not just (underpaid and unprotected) moderators directly or indirectly hired by companies. Anyone accessing and using these spaces is part of the process.

Step-by-step, Gillespie outlines a detailed and intelligent path through mechanisms of CCM gathering information from existing literature and filling the gaps in such pre-existing information enforcing his argument introducing evidence gathering news material, interviews and one-to-one conversations with the individuals shaping and enacting moderation. This is surely a necessary book to read, as the role of social media companies within social constructs becomes more and more controversial and debated. What Gillespie does, is a promising starting point to eventually access a structure that is, so far, kept out of view.

## References

- Browne, S. (2015) *Dark matters: On the surveillance of blackness*. Durham, NC, Duke University Press.
- Dawson, A. (2018) *Facebook Censors 30,000 Year-old Venus of Willendorf as ‘pornographic’*, in “The Art Newspaper”, August 09, 2018, in <https://www.theartnewspaper.com/news/facebook-censors-famous-30-000-year-old-nude-statue-as-pornographic> (retrieved March 18, 2019)
- Gillespie, T., Boczkowski, P.J. and Foot, K. A., (eds.) (2014) *Media Technologies. Essays on Communication, Materiality, and Society*, Cambridge MA, MIT Press.
- Grimmelmann, J. (2015) *The Virtues of Moderation*, in “Yale Journal of Law and Technology”, 17 (1), pp. 43-109.
- Roberts, S. T. (2016) “Commercial Content Moderation: Digital Laborers’ Dirty

- Work, in S. U. Noble and B. M., Tynes B. M. (eds.) *The Intersectional Internet: Race, Sex, Class and Culture Online*, New York, NY, Peter Lang, pp. 147-59
- Romano, A. (2018) *A New Law Intended to Curb Sex Trafficking Threatens the Future of the Internet as We Know It*, in “Vox”, July 02, 2018, in <https://www.vox.com/culture/2018/4/13/17172762/fosta-sesta-backpage-230-internet-freedom> (retrieved February 13, 2019).
- Star, S. L. (1999) *The ethnography of infrastructure*, in “American behavioural scientist”, 43 (3), pp. 377-391.
- Woolgar, S. (1990) *Configuring the user: the case of usability trials*, in “The Sociological Review”, 38 (1), pp. 58-99.

\* \* \*

### **Andreas Metzner-Szigeth (ed.)**

*Zukunftsfähige Entwicklung und generative Organisationskulturen: – Wie wir Systeme anders wahrnehmen und Veränderung gestalten können, [Sustainable development and generative organisational cultures: How we may perceive systems differently and design change]*, Munich, Oekom, 2018, pp. 256

### **Cornelius Schubert** Universität Siegen

This edited volume situates itself in the large, if not immense, domain of sustainability research and asks how change might be initiated, especially with respect to artistic interventions in organisational/entrepreneurial settings. It also focuses on a specific aspect of sustainability, namely creating novel and durable arrangements for addressing the pressing challenges of our times, by dividing the term sustainability into two components: renewability in terms of conserving resources (“Nachhaltigkeit”) and future viability in terms of creating enduring solutions (“Zukunftsfähigkeit”). In a nutshell, the book traces the possibilities of artistic interventions in order to create durable organisational changes with respect to sustainability goals such as health, equality or energy. It does so by zooming into this subject area through 15 chapters from 17 contributors. The first part of the book outlines the future viability of current societies on a broader scale. The second part questions in how far “generative organisational cultures” hold promising futures. The third part then collects examples where interventions from art or psychology have initiated dura-