Tipping the Scale: Notes on the Topologies of Big Data Platforms

MAY 2019

Platform Politick: A Series

IT FOR CHANGE

MAYA INDIRA GANESH

This report was produced as part of the research project ‘Policy frameworks for digital platforms - Moving from openness to inclusion’. The project seeks to explore and articulate institutional-legal arrangements that are adequate to a future economy that best serves the ideas of development justice. This initiative is led by IT for Change, India, and supported by the International Development Research Centre (IDRC), Canada.

Author

Maya Indira Ganesh is a technology researcher, writer and doctoral candidate at Leuphana University, Lüneburg; and is a research associate at the new research group on Artificial Intelligence and Media Philosophy at the Karlsruhe University of Art and Design. She has worked at the intersections of human rights, online activism, and digital security and privacy for over a decade.

Research coordination team

Principal Investigator: Anita Gurumurthy
Co-investigators: Deepti Bharthur, Nandini Chami
Editorial Support: Mridula Swamy, Amruta Lakhe
Design: Meenakshi Yadav, Prakriti Bakshi

Cover photo credits: Maya Indira Ganesh

© IT for Change 2019

Licensed under a Creative Commons License Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4)

IT for Change

IDRC | CRDI
International Development Research Centre
Centre de recherches pour le développement international

Canada

Tipping the Scale
Notes on the Topologies of Big Data Platforms

Maya Indira Ganesh
Leuphana University

IT for Change | 2019

Page left intentionally blank

1. Introduction

"Amazon is an elephant by design, and we're all blind men.
It may not even be an elephant; it could be five."
- Paul Ford.

"To be one, you have to be many, and that's not a metaphor."
- Donna Haraway.

We scale, or "inter-scale", by seeing fragments because the whole may be too overwhelming to consider

In September 2018 I traveled from Sydney to Singapore on a budget airline that had removed all its onboard screens. Never having visited Australia before, flying over it without a map was both intriguing and frustrating. There was no way to know exactly where we were. I did the only thing one could do in such a situation: I took photographs.

Photographs taken from inside an airplane at 30,000 feet over the Australian desert and over Greenland may be an odd starting point for an essay about big data platform infrastructures. But the view from above is deliberate: it presents a distinctive experience of scale. The vast scale of the planet is knowable and familiar thanks to a variety of representational media from in-flight maps to television programming to historic images taken from outer space. Yet, when stripped of a contextualizing reference like an in-flight map, the actual scale of the planet can be overwhelming. We scale, or "inter-scale", by seeing fragments because the whole may be too overwhelming to consider. We discursively produce scale through language-scientific and quasi-scientific systems of ordering experience into metrics; and when metrics fail, they give way to metaphors (Carr & Fisher, 2016, p. 134).

In this essay, "fictions" are proposed through which to examine the scale of platforms and the work of scale: irregular and dynamic, a series of entanglements and dependencies between human and non-human, individuals and institutions

In this essay, I propose "fictions" (Knorr Cetina, 1994) through which to examine the scale of platforms and the work of scale: irregular and dynamic, a series of entanglements and dependencies between human and non-human, individuals and institutions; not a top-down, programmed entity[i]. To see a platform in such a way is to be alert to the ways in which we might rethink how their power might be managed and contained. Unlike the wealth and power that remains concentrated at the top, as if it were a pyramidal structure, the diverse, spread-out, components comprising platforms also exert their own force within the system.

This essay is inspired by Donna Haraway's practice of "thinking-with" to make "thick gatherings" by "engaging with inherited worlds by adding layers, rather than analytical disarticulation" (cited by Puig de la Bellacasa, 2012, p.201-02). Thus, I draw on a variety of sources such as mainstream technology news and reporting, feminism, computer science, critical algorithm studies, and media and cultural studies toward theorizing what

i. "Fictionality" is deeply woven into our understanding of the scientific and technical world: "they lift modern institutions out of the purely technical but also out of the purely social and transform them into cultural systems of multiple order" (Knorr Cetina, 1994, p. 17).

anthropologist Genevieve Bell refers to as a "new kind of applied science" (2018), new "cyber-physical" systems that inhabit our world and the challenges and opportunities in their "safe, ethical and effective design, integration, management and regulation." I believe that this line of inquiry is urgent, for as Susan Schuppli (2014) writes,

> Decision-making by automated systems will produce new relations of power for which we have as yet inadequate legal frameworks or modes of political resistance – and, perhaps even more importantly, insufficient collective understanding as to how such decisions will actually be made and upon what grounds. Further, demands for public accountability will require much greater participation in the epistemological frameworks that organize and manage these new techno-social systems, and that may be a formidable challenge for all of us (p. 7).

The word platform can be misleading: we are not referring to something level, flat or raised; nor something linear, evolving in response to the formula of input → black box → output. In topological terms, we are talking about something expansive; "tentacular"; hardware and software working across different locations and temporalities; fed by a far-reaching supply chain; and, embedded in a logistical system animated by human and non-human entities

"Platform", following from its shape as something raised, level and flat, when read discursively, is an "ideological feature" (Gillespie, 2010, p. 351) as much as a physical one - "a progressive and egalitarian arrangement, promising to support those who stand on it" (p. 350). The computational, architectural, figurative and political meanings of platform as a model have been intentionally wrought to create certain discursive effects, says Gillespie. Following in this vein, I argue that the word platform can be misleading: we are not referring to something level, flat or raised; nor something linear, evolving in response to the formula of input → black box → output. In topological terms, we are talking about something expansive; "tentacular"; hardware and software working across different locations and temporalities; fed by a far-reaching supply chain; and, embedded in a logistical system animated by human and non-human entities. The focus on the topologies of platforms comes from Keller Easterling (2016), who suggests that reading platforms' topologies tells us about how information flows, and thus how complex systems are organized. The potential, or "*disposition*" (p. 76-78) for agency in platform infrastructures, "while beyond complete comprehension," lies in the relationships between things and in "activities that diverge from the stated intent" (p. 72).

It is this aspect of platforms that I explore in the essay where scale is more than just its reach in terms of numbers of users or advertising revenues. Scale is about a series of entanglements that we cannot walk away from or undo, and this makes it difficult to draw maps of the places that platforms inhabit. To scale is to have to re-draw maps and arrive at new modes of knowing and knowledge production. This essay examines scale in terms of the mathematical function of optimization which re-shapes how platforms appear and how they scale. We tend to think of scale in opposition to context, but optimization, in fact, weaponizes the notion of "context" in order to deliver goods and services calculated to be the most profitable. We tend to think of context as a kind of "stable container" (Dourish, 2004) for unique-ness, "that unquantified remainder that haunts mathematical models, making numbers that appear to be identical actually different

from each other" (Seaver 2015, p.4). I am interested in the work that scale does as the aspiration of, and as a feature of, platforms; and how we understand the implications of the constitution and shape of platforms for their management and regulation.

2. An Entwined Provenance

'Platforms' are increasingly indistinguishable from what we think of as "infrastructures," and vice versa; both have implications for governance and society. Google invests in public internet infrastructure through projects such as *Loon* [ii]. Facebook provides mobile internet in more than 40 countries worldwide through its Free Basics platform [iii]. Amazon sells books and media but also offers critical internet infrastructure [iv]. Erstwhile public services and utilities like transport and electricity in Europe and North America were splintered under Thatcher and Reagan and became privatized. Or, what used to be "television" is now streamed through platforms such as Netflix and Amazon. Tung-Hui Hu discusses how the electricity grid was used a model by scientists researching a way to deliver computing power (2017); and in *A Prehistory of the Cloud* (2015) he demonstrates how older infrastructures like television circuitry, sewers and railroads enabled the formation of the "neoliberal fantasy about user participation" that is today the cloud that powers all platforms. Platforms like Amazon, Apple, Facebook,Google or Microsoft can feel like infrastructures because they do more than just sell things and deliver search results, or because they are used by billions of people around the world. Additionally,what we tend to think of as infrastructures like city transport systems might actually be digital platforms accessed through mobile phone apps.

'Platforms' are increasingly indistinguishable from what we think of as "infrastructures," and vice versa; both have implications for governance and society

Platforms and infrastructures, though "genetically different" (Plantin et al, 2016, p. 14) are entwined in material terms. These differences include how infrastructures suggest public relations and shared and open standards, while platforms suggest closed, proprietary standards and private relations. As one takes on the dimensions of the other, their entwining contributes to platforms' (and infrastructural) complexity (p.7); a "theoretical bifocal" emerges through this: the "platformization of infrastructure" and the "infrastructuralization of platforms." It becomes increasingly difficult to consider one without the other. Therefore, software, working at scale, intersects with physical space and social geographies to actively mediate everyday life (Graham, 2005); against the backdrop of a tension between corporate control vis-a-vis the state, which, in the case of infrastructure, is a matter of public interest.

ii. "Loon LLC is an Alphabet Inc. subsidiary working on providing Internet access to rural and remote areas. The company uses high-altitude balloons placed in the stratosphere at an altitude of about 18 km (11 mi) to create an aerial wireless network with up to 4G-LTE speeds." See https://en.wikipedia.org/wiki/Loon_(company)
iii. See https://info.internet.org/en/story/free-basics-from-internet-org/
iv. Amazon Web Services (AWS) provides cloud computing infrastructure to most of the internet today and in 2017 made US$ 17.4 billion. An outage on AWS can result in vast swathes of the internet being inaccessible. See https://en.wikipedia.org/wiki/Amazon_Web_Services

Software, working at scale, intersects with physical space and social geographies to actively mediate everyday life; against the backdrop of a tension between corporate control vis-a-vis the state

Nowhere is this more apparent than in the terrain of the 'smart city'. In 2017, Sidewalk Labs, an Alphabet Inc company (formerly known as Google Inc), won a bid to partner with the city of Toronto, Canada, to redevelop its waterfront, known as Quayside, as a "smart city" project. However, it is facing resistance from citizens' groups that claim that the company's interest in environmentalism, sustainability and urban issues is in fact a Trojan horse for its surveillance capitalism [v] through mass data collection (Barth, 2018). In Berlin, Germany, a proposal to open a Google *Campus*, a space for local start-ups, was withdrawn after strong citizen protests against the increasing gentrification of the city, particularly its Kreuzberg neighborhood (Knight, 2018). These protests are against "platformization", which refers to what platforms do, what they make happen, not only what they are, that is, how they re-shape conditions of labour, economics, citizenship, commerce and geopolitics [vi]. The word platformization has been used differently in other contexts to mean programmability, (Helmond, 2015) following the origins of what a platform is: "something that can be programmed" (Andreesen quoted in Gillespie, 2010). But programmability also embodies questions of power and control manifest in the tensions between "users' goals of expression (on the one side) and, on the other side [their] profit-seeking aims and the legal surround that defines legitimate use."(van Dijck in Plantin et al 2016, p. 5). Thus, platforms do not arrive with a clean history; instead, they are entangled in other social and political forms and histories and are the sites of contestations over power and agency.

3. Planetary Scale Entanglements

Platforms do not arrive with a clean history; instead, they are entangled in other social and political forms and histories and are the sites of contestations over power and agency

The Australian desert or Greenland as seen from above are "the world without us" (Thacker, 2011, p.5) and what will remain long after we are gone. They present a complexity and unfamiliarity that are "both impersonal and horrific" (p.6). There are no neighborhoods you once lived in, towns you will ever visit, or famous landmarks to be spied from above. Yet, we are drawn repeatedly to these spaces and make an effort to know them by calling them "earth" or "the world in itself"' (p. 4). We give it meaning and want to map and study it; we want its scale to be contextualized, in order to know it better. Thinking about platforms in terms of the planet is to think about new, critical geographies of spatiality, power, international relations, the law and technology (MacDonald, 2007).

Platforms are constituted in close relationships between "platform computation" and "planetary scale computation" that generate new abstractions of platform architecture and topologies. The "planet" is invoked as a metaphor to refer to the scale of computation being

v. Surveillance capitalism as defined by scholar Shoshana Zuboff: "radically disembedded and extractive variant of information capitalism" based on the commodification of "reality" and its transformation into behavioral data for analysis and sales." https://en.wikipedia.org/wiki/Surveillance_capitalism

vi. A definition proposed by Anita Gurumurthy at the Platforms: Towards Inclusion and Openness research meeting in Bombay, June 14-15, 2018

undertaken in making complex models and calculations: "irregular systems with behaviors that are inherently discontinuous and cannot be described by equations. Such systems include the Internet, vehicular traffic flow, financial markets, and the evolution of viruses" (Lawrence Livermore National Laboratory, 2014).

The planet is quite literally implicated in the scale of platforms. For example, platforms like Facebook and Amazon run on data centers that require the scale of the planet to function: data centers are cooled by seawater (Korosec, 2011) and the planet is "terraformed" by the extraction of precious minerals to make laptops and mobile phones (Bratton, 2014).

The planet is quite literally implicated in the scale of platforms. For example, platforms like Facebook and Amazon run on data centers that quite literally require the scale of the planet to function: data centers are cooled by seawater and the planet is "terraformed" by the extraction of precious minerals to make laptops and mobile phones

Planetary scale computation refers to how the planet is sensed and measured through computational methods, from satellites to sensors embedded in trees and the soil (Gabrys, 2016). These sensors "are also collecting data on any number of environmental processes that include managing cities and facilitating logistics, as well as providing and harvesting a range of data to and from smartphone users" (p.3). This environmental data is integrated into 'smart city' platforms. In a sense, the planet itself acts as a data source through which other things might be programmed.

Thus, interpreting platforms in terms of a planet and vice versa suggests scales of actors and agencies that de-center human labor, values and creativity. Platforms comprise human and non-human agents including: individuals and organizations that imagine, design, build and maintain the computation in platforms; the imaginaries of what these platforms will do and various media industries that shape and promote these narratives; the mineral basis of hardware infrastructure from data centers to laptops and servers; the supply chains, suppliers, factories and workers that carry these minerals and metals to factories where hardware is built; millions and billions of unique users, programmers, bots, hackers, and content moderators; workers and managers; data auctions; the statistical, financial models that ascertain and project platforms' values; financial institutions; office spaces, interfaces, design, user experience; market research; and the big data and machine learning processes taking place across this hardware infrastructure. But each of these is also its own set of dense layers that become little universes in themselves. Thus, a platform at scale is a dizzying entanglement of humans, non-humans and post-humans working together to create the material embodiment of a platform that a user experiences on a device such as the mobile phone.

Such an approach to mapping entangled actors acknowledges the dense materiality of platform infrastructures, a project taken on by *Anatomy Of An AI System* [vii] (Crawford & Joler, 2018) for example. *Anatomy Of An AI System* addresses the scale of a platform through a critical presentation of

vii. See https://anatomyof.ai

its materiality. Unpacking Amazon's Echo device in terms of its planetary-scale of "extractivism", the project establishes a particular modality for thinking about the topologies of platforms. It unpacks the layers of "deep interconnections between the literal hollowing out of the materials of the earth and biosphere, and the data capture and monetization of human practices of communication and sociality in AI"(ibid, n.p) that are compressed and made invisible in the Echo's smooth, cylindrical morphology. From the colonial history of rubber tapping, to the extraction of minerals from the earth, to a theoretical framing of labor, value extraction and production, shipping and "container-ization", *Anatomy* suggests that to understand big data platforms means to entirely re-think the architecture of these systems.

The idea that breakdowns and errors on big data platforms - hacking, data breaches and privacy violations - can be traced back to a single individual or specific process is in direct opposition to the notion that platforms are complex, socio-technical, human and non-human

Thus "new ethicopolitical possibilities" (MacDonald, 2007, p. 603) required to hold these systems to account means to think beyond data privacy or surveillance by the Echo, to understanding how platform technologies are implicated in exploitative labor practices as well as exploitative environmental practices. Acknowledging the planetary scale of platform-infrastructural computation is to take the most mundane, banal and small relationships between things seriously (Haraway, 2003) and this is also an act of scaling.

4. Scale as Re-Drawing the Map

If a platform is, in fact, stacks of dense entanglements it is impossible to arrive at a conclusive map of what a platform actually looks like. Scaling goes beyond mapping and familiar visual-representational media, as well as the quite literal invocation of the globe as the aspirational scale of many platforms. For example, Facebook's ambition for planetary scale coverage is exemplified through the world map of the "global community" that Mark Zuckerberg says he is committed to building and growing (Zuckerman, 2017). Logging in to a platform from a smartphone and ordering a service, such as transportation or flower delivery, we are usually unaware of the complexity of who and what vital beings and objects are engaged in the production of those services (Hodder, 2014, p. 20). Arguing that platforms are distributed, irregular and hydra-headed has implications for how platforms evolve and are maintained managed and regulated. The view of platforms as entanglements implies that "entanglements gradually increase in complexity and scale, and it becomes more and more difficult to turn back" (p. 31).

The idea that breakdowns and errors on big data platforms - hacking, data breaches and privacy violations - can be traced back to a single individual or specific process is in direct opposition to the notion that platforms are complex, socio-technical, human and non-human assemblages. Consider the case of providing ethics education to computer scientists to address algorithmic bias [viii]. There is a belief that if computer scientists had more awareness of the social impact of their work, they might ask more probing questions before they design and develop platforms. This, however, only

serves to highlight one set of powerful actors. Some might argue that for a computer scientist to work ethically in an unequal, extractivist, capitalist system, is impossible. The unequal conditions under which Chinese workers at Foxconn assemble high-end Apple phones and laptops, for example, are sustained by complex geopolitical, financial, commercial and legal practices and histories. Changing these conditions would perhaps involve developing new kinds of practices that reach across all these systems and their interconnections. The producer of the Fairphone is taking this seriously, with the tag line "you shouldn't have to choose between a great phone and an ethical supply chain" (Fairphone, 2018). [ix] Fairphone builds its mobile phone in a radically different way and thus establishes new kinds of entanglements and networks.

Vast and complex, the value of platforms is contingent on the promise of scale, and in turn enables scaling as these concatenations of human and non-human are strung together, reaching around the planet. But, it is scale itself that unravels scale as the promise of an always-available infrastructure. So an automated system for monitoring Facebook statuses might work at scale but in the process, context is either lost, erased or just ignored (boyd & Crawford, 2012). Increasingly, "industrial" approaches to content moderation (Caplan, 2018) use human and automated content moderation, the latter running on machine learning. The writer Tom Egeland, posted a series of pictures on Facebook that he thought had "changed the history of warfare." One of these was the iconic 1972 photograph by Nick Ut of Kim Phuc, a young girl who was running away from a napalm attack during the Vietnam war. Facebook removed the picture and prevented the writer from posting on the platform (Hansen, 2016). To the algorithm managing Facebook's newsfeed, the picture was tantamount to child pornography, which was verified by a human content moderator (Scott & Isaac, 2016). For platforms working at scale, the automated screening of content is the future, and as Mark Zuckerberg has said, "over the long term, building AI tools is going to be the scalable way to identify and root out most of this harmful content" (The Washington Post, 2018).

Vast and complex, the value of platforms is contingent on the promise of scale, and in turn enables scaling as these concatenations of human and non-human are strung together, reaching around the planet

For platforms working at scale, the automated screening of content is the future, and as Mark Zuckerberg has said, "over the long term, building AI tools is going to be the scalable way to identify and root out most of this harmful content"

However, despite the likely increase in the use of automated software for content moderation, human moderators continue being employed for the task. Two recent documentaries, *The Moderators* (Cassidy & Chen, 2017) and *The Cleaners* (Block & Riesewieck, 2018) have demonstrated the human, material and affective infrastructures of 'industrial scale' (Caplan, 2018) content moderation being done in cities like Bombay and Manila on behalf of platforms headquartered in California and being used all over

viii. In 2018, the Mozilla Foundation announced a "responsible computer science challenge" to "suppor[ting] the conceptualization, development, and piloting of curricula that integrate ethics with undergraduate computer science training, educating a new wave of engineers who bring holistic thinking to the design of technology products."
See https://foundation.mozilla.org/en/initiatives/responsible-cs/
ix. See https://www.fairphone.com/en/

We have to ask, at this moment when automated content moderation practices have been shown to be so fallible, why it is that we have quickly normalized scale as speed, profit and volume

the world. While the company, say Facebook, issues them content moderation guidelines, moderators also have to exert their own judgment on the content they are evaluating. Both films demonstrate the specific contexts in which the human content moderators work, as well as the effects on them of doing this work. The *Cleaners* is a difficult film, for it shows how the task of cleaning the internet of abuse, beheadings and sexual violence takes a serious emotional toll on the young people employed to do this. Maps of platforms at scale do not capture the human, affective component of these infrastructures.

The challenges of scaling have come into sharp focus in the case of the Facebook-enabled Rohingya genocide in Myanmar. A Reuters investigation (Stecklow, 2018) has shown that despite repeated warnings from scholars and activists working in Myanmar, Facebook did not assign resources to address the incitement of violence against the Rohingya people that was being shared through Facebook posts. Facebook is the largest and most dominant news and media platform in Myanmar; it is, in fact, "the internet". Its automated content moderation system did not pick up that the derogatory slur, "Kalar", was repeatedly used online. Such a term would only be understood by someone from, or familiar with the context of Myanmar, yet, Facebook had merely 60 people with knowledge of Burmese to moderate content being shared by 15 million active users.

These instances of problematic content moderation practices on Facebook underscore the place of machine learning in big data platforms. Very simply put, machine learning is a process by which algorithms follow certain rules in order to identify patterns and associations between data points in a data set, and then use this "learning" about how things are connected to analyze new data sets. Errors are likely to be high for a number of reasons, one of which is that if a data set that a machine learning system is trained on does not contain particular contextual realities - such as words like Kalar. Human verbal language and expression is one of the more challenging tasks for machine learning. For a computer system to learn subtleties and nuances in speech and the historical context and significance of content - for example, not every picture of a naked girl is child pornography - is a little like making a 1:1 map, or a map the size of the territory being mapped. Scale, then, is to produce the world anew to be intelligible to a machine system. We have to ask, at this moment when automated content moderation practices have been shown to be so fallible, why it is that we have quickly normalized scale as speed, profit and volume.

Context is often invoked as a response to the frictions caused by multi-scalar relationships. If only something were contextual, or contextualized to the needs of a particular community or language group, there might be a more relevant and more sensitive application of a technology or platform

5. Scale as Optimization

Language is all about context. Context is often invoked as a response to the frictions caused by multi-scalar relationships. If only something were contextual, or contextualized to the needs of a particular community or language group, there might be a more relevant and more sensitive application of a technology or platform. From explicitly humanistic or social-scientific perspectives, we often argue *for* context-having it is good and taking it out is bad-although, what good and bad and context itself

mean need to be contextualized (Seaver, 2015, p. 5). Seaver cites Dourish's research to show that context, from a social-scientific perspective, tends to be "interactional" ("relational properties occasioned through association"), but in a machine system is representational ("a stable container for activity") - and that the two are incompatible. Thus, from a social science or humanistic perspective, context is not the same thing that is being modeled in a computational system (ibid).

To a machine learning system inside a big data platform, context means something entirely different. Context is central to the functioning of data-based personalization and recommender systems like Spotify, Netflix or Amazon. These systems are highly attentive to the context of our use of platform-based services and offer options to read, listen to, watch or buy depending on how our individual patterns of consumption are recorded and analyzed. When it comes to personalization and micro-targeting of individuals as consumers, this kind of context is considered "bad" or "creepy" as it violates data protection and privacy. However, this sort of "contextualization" underpins optimization, a key factor that enables platforms to scale, and ultimately, to be profitable. Optimization is the mathematical function that can be thought of as a link between context and scale. Overdorf et al (2018) write that "optimization systems treat the world not as a static place to be known, but as one to sense and co-create, poses social risks and harms such as social sorting, mass manipulation, asymmetrical concentration of resources, majority dominance and minority erasure"(p.1). "Search Engine Optimization" or SEO, is perhaps the most prevalent kind of internet marketing strategy, and is a kind of optimization at work. It is a process by which web content can become more visible by analyzing how search algorithms respond to that content and what search strings deliver the most content on a topic.

Optimization is at work when Uber identifies popular routes - say, from the city center to a highly populated suburb at 7.30am, 3pm and 6pm - then pushes that route to the top of a list that a driver will view, and then adds an extra price to it, but with meager rewards to the drivers

Optimization is at work when Uber identifies popular routes - say, from the city center to a highly populated suburb at 7.30am, 3pm and 6pm - then pushes that route to the top of a list that a driver will view, and then adds an extra price to it (the "surge"), but with meager rewards to the drivers. Certain routes, search results, status updates, hashtags, or prices for goods and services we seek out on platforms are optimized (by machine learning algorithms) in order to either become more visible and accessible, or to become less visible and accessible. 'Context' is when an optimization function shows you more advertising for products, statuses, events, routes or flights in response to [what the machine learning system thinks is] who you are online, and whom you resemble. Optimization responds to an analysis of what might return the highest profits. So, context is not about equality of opportunity or recognition, but eventually about being manipulated in the service of profitability.

Overdorf and her colleagues write that optimization techniques result in negative effects, or "externalities" (2018, p. 5), that might range from creating subsets of "high value" users, which also results in some users being considered as marginalized or "low value"; experimental testing on

Optimization is possible because of the large scale, constant collection of user data

users to explore risks and negative outcomes; "reward hacking", or developing shortcuts to achieve an optimization goal quickly; and eventually, mass data collection. Optimization is possible because of the large scale, constant collection of user data. They write that one of optimization's dominant externalities is discrimination: "Optimization systems built on data from a particular area or "domain" may underperform or downright flounder when deployed in a different environment, e.g., a voice recognition algorithm that is only trained on men's voices fails to recognize women's voices" (p. 5).

Older histories of discrimination are entangled with the machine learning systems being used to optimize platforms. Consider the case of digital redlining in the United States and Canada that follows the historic practice of redlining. According to Wikipedia, redlining was a practice from the 1930s of drawing red or pink lines on city maps to denote where lower-income and communities of color predominantly lived, thus identifying them as unsuitable for financial and economic development activities[x]. While redlining practices are against the law now, digital redlining persists as its new avatar. For example, in 2016, a Bloomberg investigation found that Amazon Prime's same day delivery service, was not available in select US zip codes in several cities like Atlanta, Boston, Chicago, Dallas, New York City and Washington DC (Ingold & Soper, 2016). These zip codes corresponded with predominantly black neighborhoods. This was stark in the case of the city of Boston, where the map of same day delivery shows the entire city as eligible for the service with the exception of Roxbury, a small pocket right in the middle, which is predominantly black. The rationale from Amazon was that a "solely data-driven calculation" and "the math involved" determined that it did not justify the expense of same day deliveries to certain parts of the city (ibid, n.p.). It is likely that there was no active discrimination by individual executives at Amazon. Instead, an optimization function may have made the decision; it might have been constructed to determine the locations where Amazon customers deliver the most value - most likely predominantly white neighborhoods, where people tend to be more wealthy.

Optimization perpetuates discrimination because of how older histories of discrimination are entangled with the machine learning systems being used in platforms

6. Epilogue: Seeing Scale as Irregularity

Shopping on Amazon from a mobile phone interface, and having deliveries made a few days later, makes the platform appear as unified and linear; but it is not. As this essay has shown, thinking about a big data platform in topological, cultural, planetary and spatial terms allows us to think about "new qualities...conventions, techniques, forms, genres, concepts … even senses" (MacDonald, 2007, p. 603). The scale at which platforms work does not suggest coherence or smooth surfaces, but irregularity (Fisch, 2018), uncertainty, instability and aberrations; and these become more apparent when platforms break down, cause harm or perpetuate discrimination. Coupled with their wide variety of uses and functions, this makes platforms difficult to define and regulate (Gillespie, 2018). So what

x. See https://en.wikipedia.org/wiki/Digital_redlining#cite_note-17

might some new "ethicopolitical possibilities" (MacDonald, 2007, p. 603) for this terrain be?

One might be the workers' resistance movements in Silicon Valley. These movements are protesting racial, class and gender disparities in hiring practices in big tech companies (Paulas, 2018); the covering up of sexual harassment and exploitation of women employees (Wakabayashi and Benner, 2018); and their employers building technologies used to perpetuate discrimination and bring harm to already-marginalized communities (Captain, 2018). And to extend the idea that platforms are comprised of myriad, diverse actors: in addition to workers, there are share-holders, advisors, investors, and everyone else enjoying the wealth and power of platforms. Where are their acts of resistance and reflection?

Another might be initiatives such as the General Data Protection Regulation, or GDPR. Despite its challenges, the GDPR essentially calls for a re-architecting of how platforms function and scale through the regulation of data flows.

Overdorf et al (2018; 2019) are researching and developing approaches to building Protective Optimization Technologies (POTS) to mitigate the effects of mathematical optimization from outside a computational system, and without the involvement of Optimization System Providers (OSPs), in other words, platforms [xi]. In an introduction to their workshop, 'a revolutionary guide to counter-optimization', they write: "Protective Optimization Technologies (POTs) enable optimization subjects to autonomously or collectively defend against negative consequences of optimization systems without having to rely on service providers." (Overdorf et al, 2019). In this workshop, Overdorf and Kulynych engaged a group in thinking through optimization systems and their 'externalities', and what kinds of POTS people might suggest as a counter-technique from outside the system (ibid). Some responses were to use privacy-enhancing tactics to disrupt the influence metrics of social media; organize movements of gig workers; disrupt the unstable, 'passionate relationship' culture of online dating; and so on."

Worker movements, POTS, and the GDPR's restrictions work at very different scales and spatialities; Worker movements are local, POTS are speculative, and the GDPR attempts to match the scale of platforms. All attempt to draw attention to what is broken and limited in the sociality, culture and governance wrought by big data platforms. They expose and resist these points of broken-ness, and work off from platform irregularities, rather than enhance the myth of smoothness. Like these, other initiatives might start to approach platforms in terms of the aberrant practices of citizenship, governance, politics and economics they create.

The scale at which platforms work does not suggest coherence or smooth surfaces, but irregularity (Fisch, 2018), uncertainty, instability and aberrations; and these become more apparent when platforms break down, or cause harm or perpetuate

xi. Further reading about the project: https://www.esat.kuleuven.be/cosic/questioning-the-assumptions-behind-fairness-solutions/

References

Bell, G. (2018). Keynote presentation at the Annual Conference for the Society for the Social Studies of Science, Sydney, Australia, August 29-September 1.

Block, H., & Riesewieck, M. (2018). *The Cleaners* [Documentary]. Germany: Gebrueder Beetz Filmproduktion.

Bratton, B. (2014). *The Stack: On Software and Sovereignty.* Cambridge and London: MIT Press

Barth, B. (2018, August 8) The fight against Google's smart city. *The Washington Post*. Retrieved from: https://www.washingtonpost.com/news/theworldpost/wp/2018/08/08/sidewalk-labs/?noredirect=on&utm_term=.f456dbf0d488

boyd, d. & Crawford, K (2011) Six Provocations for Big Data. A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society. doi: http://dx.doi.org/10.2139/ssrn.1926431

Caplan, R. (2018) Context or moderation: Artisanal, Community-Reliant, and Industrial Approaches. *Data & Society Research Institute.* Retrieved from: https://datasociety.net/output/content-or-context-moderation/

Captain, S. (2018, July 17). Meet the Silicon Valley socialists who are pushing a tech worker uprising. *Fast Company.* Retrieved from: https://www.fastcompany.com/90202378/meet-the-silicon-valley-socialists-pushing-a-surge-in-tech-worker-activism

Carr, E., & Fisher, B. (2016). Interscaling awe, de-escalating disaster. In Carr E. & Lempert M. (Eds.), *Scale: Discourse and Dimensions of Social Life* (pp. 133-156). Oakland, California: University of California Press. doi: http://doi.org/lo.1525/luminos.15

Cassidy, C. & Chen, A. (2017). *The Moderators* [Documentary]. Ireland: Field of Vision (II). Retrieved from: https://fieldofvision.org/the-moderators

Cetina, K. K.(1994). Primitive classification and postmodernity: Towards a sociological notion of fiction. *Theory, Culture and Society*. Vol 11, 1-22.

Crawford, K. & Joler, V. (2018, September 7) Anatomy of an AI System: The Amazon Echo As An Anatomical Map of Human Labor, Data and Planetary Resources. *AI Now Institute and Share Lab*. Retrieved from: https://anatomyof.ai

Dourish P (2004) What we talk about when we talk about context. *Personal and Ubiquitous Computing,* 8(1): 19–30.

Easterling, K. (2016) *Extrastatecraft: The power of infrastructure space.* London and New York: Verso.

Ford, P. (2013, March 1) The Fog Kingdom: The useful imponderability of Amazon, America's new most reputable company. *New York Magazine*. Retrieved from http://nymag.com/news/intelligencer/amazon-2013-3/

Fisch, M. (2018) Meditations on the "Unimaginable" (soteigai). Forthcoming in *The Space of Disaster Golani Solomon*, E. (Ed). Tel-Aviv: Resling Publishing

Gabrys, J. (2016). Program earth: Environmental sensing technology and the making of a computational planet. *Electronic Mediations 49*. Minneapolis and London: University of Minnesota Press.

Gillespie, T. (2010) The politics of "platforms." *New Media & Society,* 12(3): 347–364.

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

Graham, S.D.N. (2005). Software-sorted geographies. *Progress in Human Geography,* 29 (5), 562–580.

Hansen, E. E. (2016, September 8). Dear Mark. I am writing this to inform you that I shall not comply with

your requirement to remove this picture. *Aftenposten*. Retrieved from https://www.aftenposten.no/meninger/kommentar/i/G892Q/Dear-Mark-I-am-writing-this-to-inform-you-that-I-shall-not-comply-with-your-requirement-to-remove-this-picture

Helmond, A. (2015) The platformization of the web: making web data platform ready. *Social Media+ Society*, 1(2). doi: 10.1177/2056305115603080

Haraway, D. (2003). *The Companion Species Manifesto: Dogs, People, and Significant Otherness.* Chicago: Prickly Paradigm Press.

Haraway, D. (2016). *Staying with the Trouble: Making Kin in the Chthulucene*. Durham: Duke University Press.

Hodder, I. (2014). The Entanglements of Humans and Things: A Long-Term View. *New Literary History*, 45: 19–36.

Hu, T-H. (2015). *A Pre-history of the cloud*. London and Cambridge: MIT Press

Hu, T-H. (2017). Black boxes and green lights: Media, infrastructure, and the future at any cost. *English Language Notes,* 55: 1–2, 81-88. Retrieved from https://umich.app.box.com/s/ky5b6b3vypgmq9iwskzj6oj2jjjikmma

Ingold, D. & Soper, S (2016, April 21) Amazon Doesn't Consider the Race of Its Customers. Should It? *Bloomberg*. Retrieved from https://www.bloomberg.com/graphics/2016-amazon-same-day/

Korosec, K. (2011, May 24) Google's Latest Innovation: Seawater-Cooled Data Centers. *CBS News.* Retrieved from https://www.cbsnews.com/news/googles-latest-innovation-seawater-cooled-data-centers/

Knight, B. (2018, October 25) Google leaves Berlin campus after anti-gentrification demos. *Deutsche Welle*. Retrieved from https://www.dw.com/en/google-leaves-berlin-campus-after-anti-gentrification-demos/a-46045271

Lawrence Livermore National Laboratory (2014, July). Visualizing Complex Networks on a Planetary Scale. Retrieved from https://computation.llnl.gov/newsroom/visualizing-complex-networks-planetary-scale

Macdonald, F. (2007) Anti-Astropolitik: Outer space and the orbit of geography. *Progress in Human Geography*, 31(5), 592–615. doi: 10.1177/0309132507081492.

Overdorf, R., Kulynych, B., Balsa, E., Troncoso, C. & Gürses, S. (2018). POTs: Protective Optimization Technologies. Retrieved from https://arxiv.org/abs/1806.02711

Overdorf, R., Kulynych, B., Balsa, E., Troncoso, C. & Gürses, S. (2019). A revolutionary guide to counter-optimization. Workshop at the Internet Freedom Festival. Valencia, Spain. Workshop details: https://platform.internetfreedomfestival.org/en/IFF2019/public/schedule/custom/1028 Workshop notes: https://pad.internetfreedomfestival.org/p/1028

Paulas, R. (2018, September 04) A New Kind of Labor Movement in Silicon Valley. *The Atlantic*. Retrieved from https://www.theatlantic.com/technology/archive/2018/09/tech-labor-movement/567808/

Plantin, J., Lagos, C., Edwards, P.N. & Christian Sandvig (2016). Infrastructure studies meet platform studies in the age of Google and Facebook. *New Media & Society*, 1-18, doi: 10.1177/1461444816661553

Puig de la Bellacasa, M. (2012, May): 'Nothing comes without its world' : Thinking with care. *Sociological Review.* doi: 10.1111/j.1467-954X.2012.02070.

Seaver, N (2015). The nice thing about context is that everyone has it. *Media, Culture & Society, Crosscurrents Special Section: Debating Big Data*, 1-9. DOI: 10.1177/0163443715594102

Schuppli, S. (2012) Deadly algorithms: Can legal codes hold software accountable for code that kills? *Radical Philosophy*, 187 (Sept/Oct 2014), 2-8.

Scott, M. And Isaac, M. (2016, September 9) Facebook Restores Iconic Vietnam War Photo It Censored for Nudity. *The New York Times*. Retrieved from https://www.nytimes.com/2016/09/10/technology/facebook-vietnam-war-photo-nudity.html

Stecklow, S. (2018, August 15) Why Facebook is losing the war on hate speech in Myanmar. *Reuters.* Retrieved from https://www.reuters.com/investigates/special-report/myanmar-facebook-hate/

Thacker, E. (2011) *In the Dust of This Planet: Horror of Philosophy vol. 1*. Zero Books.

Wakabayashi, D., and Benner, K. (2018, 25 October) How Google Protected Andy Rubin, the 'Father of Android'. *The New York Times.* Retrieved from https://www.nytimes.com/2018/10/25/technology/google-sexual-harassment-andy-rubin.html

Wark, M. (2016, December) The Stack to Come: On Benjamin Bratton's The Stack. *Public Seminar*. Retrieved from http://www.publicseminar.org/2016/12/stack/

Washington Post (2018, April 10) Transcript of Mark Zuckberg's Senate Hearing. *The Washington Post.* Retrieved from https://www.washingtonpost.com/news/the-switch/wp/2018/04/10/transcript-of-mark-zuckerbergs-senate-hearing/?utm_term=.517903d11938

Zuckerman, E. (2017). Zuckerberg and the Imaginary Cosmopolitan. The California Review of Images and Mark Zuckerberg Vol 1, Winter 2017. Retrieved from http://zuckerbergreview.com/zuckerman.html