# CHAPTER 3. Science and colonialism: The violence of abstraction

### Teresa Numerico

## Summary

* Throughout modernity, both science and colonialism have relied on practices of abstraction that underlie both the idea of objectivity and specific practices like cybernetics and information science
* Data colonialism continues this process by automating practices of social classification, whose roots go back much earlier in modernity but which have been intensified by the forces of capitalism
* The asymmetries, biases and limits of AI and Big Data all derive from these problems of abstraction and standardization, which must be replaced with a truly relational view of knowledge

Colonialism and racial capitalism don’t change the world through brute power alone. They work through knowledge and imagination. Knowledge and imagination are key elements in how economies and societies get built. A crucial tool in the emergence of colonial and capitalist economies and societies was (and still is) *science.* This chapter explains the crucial role, behind the scenes of everyday life, that the institution of modern science has played in making colonialism and capitalism possible.

From the start of Western science, the idea of scientific progress is related to the quantification and measurement of phenomena which are the object of interest of the scientists. According to Hannah Arendt in *The Human Condition*, for the scientist to measure and conceive the scientific object they have to exit from the space of representation.[[1]](#footnote-1) This means they have to stop relying on their structure of everyday perception and trust artificial tools. The emergence of Western science thus started with Galileo’s and Descartes’ work on the mathematization of the understanding of physical phenomena (Galileo) and the measurement of space, via its geometrization (Descartes)—a work accomplished with the advent of Enlightenment, around the 17th century.[[2]](#footnote-2),[[3]](#footnote-3)

The very concept of objectivity was defined a bit later, from the start of the 18th century.[[4]](#footnote-4) However, once this concept prevailed among others it defined the idealization of the object that needed to be understood and described by the scientific method. The notion of objectivity is crucial to understand all the forms of abstracted knowledge on which modern science depends.

Contemporary ideology toward scientific method and organization of information and knowledge has many different sources. One of the most effective and influential was cybernetics, which was officially born in 1948 with a book by Norbert Wiener, entitled *Cybernetics: or control and communication in the animal and the machine*. This trans-discipline, which contributed to ideas such as Artificial Intelligence and the simulation of human behavior with machines, conceived of interaction between human beings as a form of communication. It suggested that communication was a general activity that could belong to humans, but also across different animals and machines.[[5]](#footnote-5)

Cybernetics consisted in understanding the feedback loops that were necessary to adjust conducts in relation to the environment. The idea was that these feedbacks were very common in natural contexts as well as in interacting with machines such as radars and thermostats. The mechanisms that allow animals to survive in natural hostile situations, according to Wiener, were reproducible patterns that could be simulated in machine interactions. Control was a special type of communication in which we want to make sure that the receiver agent of the message accomplished the order prescribed in the communication stance. The linguistic context of Cybernetics as well as most of the content that was its object of study was related to the Second World War and retained most of the war-like atmosphere in which it was conceived.[[6]](#footnote-6)

This approach opened science to a completely new vision of what it meant to offer a scientific explanation of a phenomenon, whose objectivity was guaranteed more by the technical system of automating data gathering and the subsequent data retrieval and sorting out in relevant context. The human role in science was limited to the governing of the general process, without any control on the analysis and the organization of data and relevant models of explanation.[[7]](#footnote-7)

Cybernetics was preceded by a seminal paper by Rosenblueth and Wiener on the role of models in science, according to which the aim and the organization of science involved a distinctive epistemic frame.[[8]](#footnote-8) What did it mean to create a scientific model? Models could be abstract or material but they both required sorting out some closed-box problems, deciding which of the variables that contributed to the phenomenon were worth measurement and which should remain hidden inside the box.

The evolution of the scientific framework of Western science was then based on the measurability, objectivity, and abstraction of relevant characteristics. The idea behind this approach to science was that the scientist should be in the universal and absolute position to judge and classify phenomena according to a quantitative, explicit, and rigorous way.

This is the scenario against the background of which the process of colonial datafication and extraction of information from human beings and human life emerges as a project that continues the broader trajectory of Western science.

## Data colonialism and the automation of classification

In modern western science there is a reliance on classification. Major efforts are made to organize knowledge in the form of a classification following a precise style of judgment that defines substances as subjects and attributes as characteristics that belong to them. Substances are hierarchically superior to attributes that are predicated of them. This is the legacy of Aristotelianism. Modern science was born against scholastic knowledge, which was supported by the Church, but though it criticized most of the fundamental premises of that epistemology, it retained the Aristotelian structure of knowledge in terms of categories, substances, and attributes.

In the western world there was also a different tradition which was based on a more relational conception of reality, both in epistemological and in ontological terms. Why did the Western tradition choose to silence, or diminish the influence of a relational approach to knowledge and understanding that would have been more inclusive, acknowledging the situatedness and subjectivity of every point of view? The answer lay in the juridical necessity to justify the appropriative attitude toward the land of indigenous people, who had to be considered less ‘human’ than the colonizers. This approach was also backed by the Christian church which was defeated by the refusal of Modern Western science to accept its authority in justifying knowledge, but rebuilt a new alliance with the scientific knowledge in the colonization of imagination.[[9]](#footnote-9)

The ideology of modern science was supported together with the first essays of capitalistic appropriation of land in the United Kingdom during the time of the British monarch, Elizabeth the First, that allowed the first enclosures of public land for private exploitation. The model of national land grabbing offered by enclosures was soon exported on the global level, due to the never-ending need of capital for new sources of appropriation to feed the infinite growth process. The UK example was followed by the recently founded nation states in Europe: Spain, Portugal and then France. There was a colonial appropriative movement also within the western world, and then it was exported abroad.

Underlying this convergence between theoretical science and practical extraction was the vision of a new relation between humanity and nature that had been announced by English philosopher Francis Bacon in his book *Novum Organon*, originally written in 1620*.* For Bacon nature existed for man to extract from it, through techniques of knowledge and force, whatever value man wanted, and without concern for the consequences.[[10]](#footnote-10)

The idea of Bacon, in fact, was not only that it was possible to extract value from land but also that this extractive process could last forever, because no resistance to it was imagined. The possibility of other human users of that same land was never considered. This landgrab was in fact the original historical accumulation which was necessary for capitalism to happen at all. According to Jason Moore in his book *Capitalism in the Web of Life*, we can read this large-scale reorganization of resources that comprised early colonialism as a search for a new frontier of primitive accumulation.[[11]](#footnote-11)

Capitalist ideology and Western science ideology could survive only if they could demonstrate the global progression of exploitation, continuous increase of resources appropriated, never-ending progress of knowledge creation and the development of technological tools that are more and more powerful.

This attitude toward unlimited appropriation together with the ideology of hierarchical categorization of scientific objects and their characteristics is not necessary for science or for knowledge. It is only necessary for capitalistic exploitation. This sets the scene for understanding the special role of *data* in contemporary science.

## Data as quantified, biased, conservative interpretation of the research object

How does this primitive accumulation connect with data? The idea that data can be a brute univocal representation of facts, without any intermediation, descends from the suggestion that phenomena can be reproduced, and eventually directly created in the form of data without the implication of a representative choice or any specific view. It presupposes that the data is not situated in a specific contextual way of representing objects. The idea of objectivity that was crucial for the idea of western modern science according to Daston and Galison relied on the invention of some technical devices that allowed the possibility to compare representations in a univocal way.[[12]](#footnote-12) In the world of data gathering the bare idea that it is possible to access such a huge amount of data counts as the accessibility of *the totality* of the object of research.

From a traditional epistemological point of view, however, this view of knowledge is completely devoid of foundation. Data is a perspective on the world, as all representations are. Data always falls short of representing the totality of what is ‘datafied’ and needs all sorts of external explanation in order to be implemented in a system whose aim is the production of future predictions, related to people's characteristics, preferences and behaviors.

We know that decision-making algorithms must use some methods/tools to interpret past data to obtain predictions of future behaviors. Many assumptions are embedded in such projections of past data onto future situations. These assumptions are often neglected, and they are not objects of attentive, explicit, conscious reflections. Such a lack of awareness together with the arrogance of the knowledge system which claims its accuracy and precision without exhibiting a correct method of collective auditing of its epistemic scaffoldings is the potential cause of the appropriative, colonizing structure of this knowledge acquisition method. This attitude is particularly dangerous because it projects its (limited) results on the future, configuring the future around the partial interpretations of past events and presumed preferences and behaviors.

There is a connection between the Western scientific approach to understanding highlighted here and racism —a connection made by Hannah Arendt.[[13]](#footnote-13) The categorial representation of the people allows easily the interpretation of difference as the indication of exclusion characteristics that are described as ‘naturally’ inferior to the proxy classification of what we are prepared to consider the standard model of subjectivities.[[14]](#footnote-14),[[15]](#footnote-15),[[16]](#footnote-16)

The infrastructure of knowledge that allows such a systemic appropriation and projection of people’s data is critical from a political as well as from an epistemological point of view.[[17]](#footnote-17) Its commercial conditions of production imply a strong asymmetry of power between those who are modeled and the subjects who are the actors of the knowledge production strategy. Data science scholars work on a technical and epistemic structure that is opaque, blurred and hidden, though its effects are rather visible on society as a whole. The asymmetry is not only a feature of data’s technical infrastructure scaffoldings but also linked to the people hired to manage data, who are all trained in the same universities and belong to a similar group in terms of gender, ethnic origin and socio-cultural milieu.[[18]](#footnote-18),[[19]](#footnote-19),[[20]](#footnote-20),[[21]](#footnote-21) This asymmetry is reinforced by how data is gathered from everyone who uses digital devices and services. Most of the people whose data is used are not themselves represented among the workers who make decisions based on that data through algorithmic processing.

In the following sections I describe some of the characteristics of the data interpretation process that deal with automated decision-making systems. The aim of the description is to suggest the following points: there is no such a thing as raw data, data are always cooked, no matter if implicitly or explicitly, there is always some model at work in interpreting data, models used in AI inferences tend to amplify past conditions to predict future events, considered inevitable and not contingent. And, for these reasons, data processing always involves a certain epistemic violence on the actual environment from which its knowledge is abstracted.

## The implicit role of context

For the correlations to make sense within the statistics of data, it is necessary that the context is clearly established, otherwise it is possible to make erroneous correlations produced by the lack of independence between data. We can, for example, erroneously deduce that people with cardiovascular problems are more protected than others against pneumonia, which is just the consequence of the fact that those patients are more monitored than the others due to special prevention measures assumed to avoid worse consequences in case of catching pneumonia. In order to understand data, then, it is necessary to have the context in which data is gathered and all sorts of co-implications of types of data.

## Quality plus abstraction produces quantification

When we use data for representing a situation and modeling it for projecting the past data in future prediction we have to choose an abstraction method in order to define how to describe the situation in quantitative format and what can be counted together.[[22]](#footnote-22) One of the crucial activities of data science is the production of clusters where data of different subjects are considered as part of a unique group. Such a process requires decisions about how to abstract between different data in order to define a common category to which they are treated as belonging.

Abstraction is a necessary activity when we want to define a proxy, for example, in an automated recruitment process. We need to decide which are the features that are considered the desired ones in order to choose a candidate for a job position. However, the definition of the model candidate is not as the definition of the model cat, because it is full of potential discriminations, depending on which characteristics we select to abstract and create the proxy of our preferred worker. The abstraction is necessary to create the profile that we want the system to learn in order to predict the most valid candidates for the shortlisting exercise. The technicality of the learning system allows us to hide the decisions inside the procedure, but the abstraction choices are there anyway and orient systemic choice in a situated way.

## Biases are necessary to the learning process

The use of machine learning methods for the creation of automated decision-making tools is based on the possibility of making sense of huge quantities of data. In order to interpret data the systems need some strategies to identify patterns that are used as examples for future recognition of similar peculiarities in testing data, once that training sets of data are identified, classified and properly tagged. The use of training sets or other relevant methods used to make sense of future data relies on some sort of learning bias.[[23]](#footnote-23) The learning bias is not necessarily negative or epistemically problematic, but it is a way to synthesize the many possible interpretations that are available to make sense of data.

The result is that if we don’t control precisely the external choices that allow the bias to happen, we risk implementing prejudices and prescriptions in the code, whose aim is to create an effective learning function to organize data with meaningful connections. The way in which we create those connections will influence the results of the system outputs. This is clearly stated in accounts of data colonialism.[[24]](#footnote-24)

One of the consequences of this learning bias over data is the tendency to project past data into future prediction. One of the implicit inferences that are adopted by machine learning methods is the use of the inductive principles to obtain sound conclusions.[[25]](#footnote-25) But, of course, if we project the past on the future the temptation is to consider the past as the measure of the future, and to normalize the conditions of the past as if they were inscribed in the nature of the subjects whose future behaviors need to be guessed in advance.

If society were a perfect system in which everybody had access to the same opportunities in life, intellectual stimuli, education and wealth, then this approach would face no fundamental objections, but the truth is that the concrete and practical conditions of human beings are *not* equal or comparably distributed. Society is unfair because of the historical unequal conditions imposed from the colonial legacy and other differences in access to privileges in wealth and cultural consciousness possibilities. If we fail to take the historical injustices into account we will repeat the inequities of the past, with the help of the oracular effect of future predictions offered by technological systems which are considered more objective and trustworthy than human predictions. Contingent conditions are absolutized by the automated abstraction process without any awareness from both who set the system in place and who use its conclusions as valid without any extra checking effort.

Data processing’s methods end up by imposing a conservative approach to predictions which are relevant to anticipate behaviors, desires, attitudes and preferences of people, inferring those judgements on their past behaviors, desires etc. Or, even worse, by clustering people in groups, associating their characteristics to belonging to the same set of people that share some common elements. These correlations that start from clusters of people to demonstrate their qualities or behaviors is particularly dangerous because it is the same attitude of traditional racism evaluations.[[26]](#footnote-26),[[27]](#footnote-27)

We cannot avoid noticing that the huge economic and technical investments behind the development of AI machine learning tools, and the tendency to amplify digital surveillance in all its possibilities,[[28]](#footnote-28) imply a deep asymmetry of power between those who face the use of their data for predicting purposes and those happy or unhappy few who are in control of the algorithmic processing of data. Understanding this context precisely means to acknowledge the fact that predictions become prescriptions, because there is no way to oppose the systemic vision imposed by AI machine learning tools.[[29]](#footnote-29) In fact, it doesn’t matter if data was not correctly *describing* the situations because the prescriptive approach to the future outcomes, together with the impossibility to enter the so-called black box where predictions get made and validated, results in a situation where the subjects of the predictions are unable to defend themselves.[[30]](#footnote-30),[[31]](#footnote-31) Data-driven predictions taken in an oracular form and there is no way to discuss them or even ask for a justification of their assumptions.[[32]](#footnote-32)

## The standardization of interpretations

How do these underlying points about the underlying nature of data processing connect with our everyday experience of the world today?

Take the example of how data systems, for example embedded within social media platforms or marketing data, attempt to define a particular set of facial characteristics as a signal that can be interpreted as a proxy for whether someone is part of a straight or gay group.[[33]](#footnote-33) The universalistic approach to knowledge as based in single ways of classifying the world is at work here. In this example such a classification is explicitly inadequate to describe the complex and highly contextual sexual identities of actual people. The boundaries between straight and gay are, after all, not the only possible distinctions between people in terms of their sexual identities.

It is not possible to think about all the possibilities of life on earth through a single classificatory grid, and this is particularly true when such classification is married with an universalistic attitude toward research and science. The risk here is the standardization of the gaze toward reality so that it is not possible to understand the nuanced possible interpretation. The standardization is prescriptive when it is applied to social and human behaviors and it renders it difficult to be accepted for people who are not perceived as normal.

However, normality is an imposition of the universalistic approach to understanding. This is yet another aspect of what Quijano called ‘the coloniality of power’ (see chapter 2 by Nai Lee Kalema). This is one of the methods of cultural imposition and appropriation that is acted by the knowledge/power structure. It is based on the asymmetry of power between who is producing the understanding process and who is the object of such a process, without any assessment, or ‘aware and informed consent’ procedure, or audit process in place. The voices of the subjects represented in the data are always muted. In fact, they are muted twice because data is always coded as silent and because there is no representation of those subjects in the processes of knowledge creation and development.

## The relational and collective approach to knowledge

It is however possible to understand phenomena from a different perspective, a pluralistic one, which is based on a relational ontology and/or epistemology as it is suggested by philosophers from Global North and Global South such as Whitehead, Stiegler, Mbembe or Yuk Hui, among others. Wendy Chun’s book, *Discriminating Data* (2021), shows that it is not necessary to interpret reality according to a unique system that is based on a precise ideology of discrimination, normalization, and standardization of human relationships and habits.[[34]](#footnote-34)

In order to work in favor of a decolonial perspective about data it is necessary to step out of the universalistic approach of western science, and to exit from modernity.[[35]](#footnote-35)

We need to provincialize Europe because there is no universal history to deploy.[[36]](#footnote-36) We have to understand to accept pluralistic interpretation of facts, habits and desires of people, so that we do not seek to hypostatize the relative digital traces transforming those traces into methods for expropriate behaviors of people by offering a univocal abstract symbolic meaning. We need to find a way to preserve digital pluriversality.[[37]](#footnote-37)

Without such a transformation of viewpoint, human beings risk being expropriated as an open terrain, and becoming a new frontier for the appropriation on which capital survives.

1. Hannah Arendt, The human condition, University of Chicago Press, 1958 (2013). [↑](#footnote-ref-1)
2. Alexandre Koyré, From the Closed World to the Infinite Universe, Baltimore: Johns Hopkins Press, 1957. [↑](#footnote-ref-2)
3. Alexandre Koyré, Metaphysics & Measurement: Essays in Scientific Revolution, Harvard University Press, 1968. [↑](#footnote-ref-3)
4. Lorraine Daston and Peter Galison, Objectivity, Princeton University Press, 2010. [↑](#footnote-ref-4)
5. Norbert Wiener, Cybernetics: or Control and Communication in the Animal and the Machine, The MIT Press, 1948 (1961). [↑](#footnote-ref-5)
6. Paul N. Edwards, The Closed World: Computers and the Politics of Dis- course in Cold War America, MIT Press, 1997. [↑](#footnote-ref-6)
7. Joseph Carl Robnett Licklider, Libraries of the Future, MIT Press, 1965. [↑](#footnote-ref-7)
8. Arturo Rosenblueth and Norbert Wiener, ‘The Role of Models in Science’, Philosophy of Science 12.4 (1945): 316–21. http://www.jstor.org/stable/184253. [↑](#footnote-ref-8)
9. Quijano, ‘Coloniality and modernity/rationality’ [↑](#footnote-ref-9)
10. Francis Bacon, Francis Bacon: the new organon, Cambridge University Press, 1620 (2000). [↑](#footnote-ref-10)
11. Jason W. Moore, Capitalism in the Web of Life: Ecology and the Accumulation of Capital, Verso Books, 2015. [↑](#footnote-ref-11)
12. Daston and Galison, Objectivity. [↑](#footnote-ref-12)
13. Hannah Arendt, The origins of totalitarianism, Houghton Mifflin Harcourt, 1973. [↑](#footnote-ref-13)
14. Allen Chun, (Post) Colonial governance in Hong Kong and Macau: a tale of two cities and regimes, Postcolonial studies 22.4 (2019): 413–427. [↑](#footnote-ref-14)
15. Virginia Eubanks, Automating inequality: How high-tech tools profile, police, and punish the poor, St. Martin's Press, 2018. [↑](#footnote-ref-15)
16. Cathy O'Neil, Weapons of math destruction: How big data increases inequality and threatens democracy, Crown Pub, 2016. [↑](#footnote-ref-16)
17. Paul N. Edwards, Geoffrey C. Bowker, Steven J. Jackson, and Robin Williams, ‘Introduction: an agenda for infrastructure studies’, Journal of the association for information systems 10.5 (2009). doi:10.17705/1jais.00200. [↑](#footnote-ref-17)
18. Jessie Daniels, ‘“My Brain Database Doesn’t See Skin Color” Color-Blind Racism in the Technology Industry and in Theorizing the Web’, American Behavioral Scientist 59.11 (2015): 1377–1393. [↑](#footnote-ref-18)
19. Sinduja Rangarajan, ‘Here’s the Clearest Picture of Silicon Valley’s Diversity yet: It’s Bad. But Some Companies Are Doing Less Bad’, Reveal, 25 June 2018. https://www.revealnews.org/article/heres-the-clearest-picture-of-silicon-valleys-diversity-yet/ [↑](#footnote-ref-19)
20. Sarah Myers West, Meredith Whittaker, and Kate Crawford, ‘Discriminating Systems: Gender, Race and Power in AI’, AI Now Institute, 1 April 2019. https://ainowinstitute.org/discriminatingsystems.html [↑](#footnote-ref-20)
21. Catherine D'ignazio and Lauren F. Klein, Data feminism. MIT Press, 2020. [↑](#footnote-ref-21)
22. Deborah A. Stone, Counting: How we use numbers to decide what matters, Liveright Publishing, 2020. [↑](#footnote-ref-22)
23. Mireille Hildebrandt, The issue of bias: the framing powers of ML. Machines we trust. Perspective on dependable AI, MIT Press, 2021.  [↑](#footnote-ref-23)
24. Couldry and Mejias, The Costs of Connection: How Data is Colonizing Everyday Life and Appropriating it for Capitalism. [↑](#footnote-ref-24)
25. Osonde A. Osoba, Benjamin Boudreaux, Jessica M. Saunders, J. Luke Irwin, Pam A. Mueller, and Samantha Cherney, Algorithmic Equity: A Framework for Social Applications, Santa Monica: RAND Corporation, 2019. [↑](#footnote-ref-25)
26. Benjamin, Race After Technology: Abolitionist tools for the new Jim code. [↑](#footnote-ref-26)
27. Wendy Hui Kyong Chun, Discriminating Data. Correlation, Neighborhoods, and the New Politics of Recognition, MIT Press, 2021. [↑](#footnote-ref-27)
28. Oscar H. Jr. Gandy, The panoptic sort: A political economy of personal information, Oxford University Press, 2021. [↑](#footnote-ref-28)
29. Antoinette Rouvroy and Thomas Berns, ‘Gouvernementalité algorithmique et perspectives d’émancipation. Le disparate comme condition d’individuation par la relation’, Réseaux 177.1 (2013): 163–196. [↑](#footnote-ref-29)
30. Frank Pasquale, The black box society: The secret algorithms that control money and information, Harvard University Press, 2015. [↑](#footnote-ref-30)
31. Frank Pasquale, New laws of robotics: defending human expertise in the age of AI, Belknap Press, 2020. [↑](#footnote-ref-31)
32. Ed Finn, What algorithms want, MIT Press, 2018. [↑](#footnote-ref-32)
33. Paul B. Preciado, ‘Dissident Interfaces: Shu Lea Cheang’s 3x3x9 and the Digital Avant-Garde’, in Preciado, Paul B. (ed.), 3x3x6. Shu Lea Cheang, Taiwan: Taipei Fine Arts Museum, 2019, pp. 69–90. [↑](#footnote-ref-33)
34. Chun, Discriminating Data. Correlation, Neighborhoods, and the New Politics of Recognition. [↑](#footnote-ref-34)
35. Yuk Hui, The question concerning technology in China: An essay in cosmotechnics: 3, MIT Press, 2019. [↑](#footnote-ref-35)
36. Dipesh Chakrabarty, Provincializing Europe, Princeton University Press, 2009. [↑](#footnote-ref-36)
37. Arturo Escobar, Designs for the pluriverse: Radical interdependence, autonomy, and the making of worlds, Duke University Press, 2018. [↑](#footnote-ref-37)