

Book Review: Data Practices: Making Up a European People by Evelyn Ruppert and Stephan Scheel

In Data Practices: Making Up a European People – available [open access](#) – Evelyn Ruppert and Stephan Scheel explore how statisticians and policymakers use statistical methods and data practices to ‘enact’ or ‘make up’ their data subjects: in this case, the people of Europe. The book’s detailed case studies and thoughtful consideration of quantitative data production from the perspective of the data subject have earned it pride of place on the bookshelf of reviewer **Mariele McKone Leonard**.

***Data Practices: Making Up a European People*. Evelyn Ruppert and Stephan Scheel. Goldsmiths Press. London. 2021.**

Statistics are often regarded in one of two ways: [as a realistic, but passive, description of the world](#) (for instance, ‘descriptive statistics’) or of a piece with lies and damn lies. Proponents of the latter view are somewhat more accurate in perceiving the effective purpose and power of statistics, if not the intent, because they capture the active nature of statistical enumeration. As Evelyn Ruppert, Stephan Scheel and their co-authors deftly show in [Data Practices: Making Up a European People](#), the production of statistics is hardly passive.

Data Practices is the result of the ‘[Peopling Europe: How Data Make a People](#)’ (ARITHMUS) project, supported by Goldsmiths University and the European Research Council. The ARITHMUS project seeks to answer the question: who are the people of Europe? As the authors show, this is really a question of how do EU statisticians and policymakers ‘[grapple with harmonising and standardising enumeration methods and data across member states to make one European population](#)’?

ARITHMUS, and *Data Practices* as a result, takes as its foundational principle that statistical methods are ‘performative’, meaning they help ‘enact’ or ‘make up’ their data subjects, in this case the people(s) of Europe (4). The book thus focuses on how statisticians and policymakers use the production of statistics, via data practices, to shape these people(s). That statistics, especially official statistics and research derived from them, are inherently political is not a new argument (see [James C. Scott 1999](#); [David I. Kertzer and Dominique Arel 2009](#)). However, unlike earlier works, *Data Practices* is less theory-driven (and thus less heavy with references to Michel Foucault and Bruno Latour) and more practicable. This will likely make it more accessible to many statisticians and data scientists who are trained in mathematics or computer science, but not in social science (although they should be).

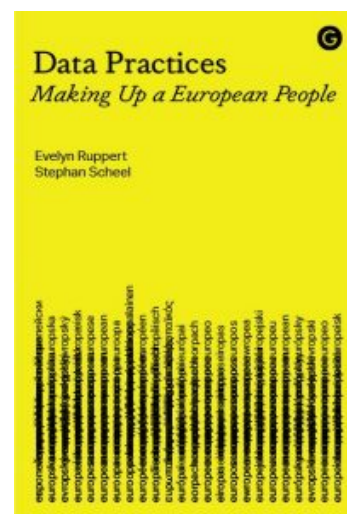




Image Credit: [Pixabay CCO](#)

In the second chapter, Ruppert and Scheel explain the term 'data practices'. Data practices are, in essence, the actions taken to generate, process, analyse and share data. Twelve of these practices are expanded upon in six subsequent chapters: defining and deriving (Chapter Three); coordinating and narrating (Chapter Four); omitting and recalibrating (Chapter Five); inferring and assigning (Chapter Six); calibrating and sieving (Chapter Seven); and differentiating and defending (Chapter Eight). Each of these chapters is illustrated by discussion of a group or groups of statistical subjects in turn: so-called 'usual residents', refugees and homeless people, migrants, foreigners, data subjects more broadly and statisticians themselves.

Through the course of the substantive chapters and their case studies, and by making explicit the theoretical and analytical decisions statistical data practices require, the authors expose the myth of statistics as passive, reflecting, measuring and representing an already existent reality: a phenomenon Morgane Labbé (2000) refers to as 'statistical realism'. This myth presents the data practitioner as a dispassionate, neutral transcriptionist of data, and treats data practices as 'simply what practitioners do' (32). This allows statisticians, data scientists, politicians and policymakers to pretend that the data 'speak for themselves [...] free of human bias or framing' ([Rob Kitchin 2014, 5](#), quoted on page 30).

The perpetuation of this myth via statistical and data science curriculums does a radical disservice to the profession, as it limits statistics and the social sciences to the superficial, and it leaves data practitioners unprepared to critically interrogate the question of why the data says what it says. After all, if the data is simply the presentation of things 'as they are', then [there is no place for bias, racism or even misuse of statistics](#).

The reality is that the history of statistics and state enumeration is fraught with racism, ableism and other forms of discrimination. Scientists and researchers have always used statistics to separate, segregate and degrade. In addition to the examples of Nazi Germany and apartheid South Africa, [Melissa Nobles](#) has documented how data on 'mulattoes' was collected in the US Census to 'prove that mulattoes lived shorter lives, and thus that blacks and whites were different racial species' (53). In a more contemporary example, the Body Mass Index (BMI) is still used as an indicator of good health, [despite its origin as a mean of the population of nineteenth-century European men, and overwhelming evidence that it is not valid for people of colour](#).

Although the authors touch only briefly on these histories and their role in shaping official definitions and policies, (see Chapter Six, 'Foreigners', especially), I don't consider this a major weakness of the work, as these topics have been covered elsewhere (see, for example, the excellent [Thicker than Blood by Tukufu Zuberi](#)). Instead, and perhaps what makes *Data Practices* most relevant today, is the considered discussion of the relative power of both data practitioners and data subjects in Chapters Seven and Eight.

As a piece of the process of rendering legible the variety of human experiences, the practice of statistics objectifies data subjects – who are, at least in the grammatical sense of the word, actors – into mere representations through extraction, simplification and modification of their data. In order to do so, any data – or persons – who cannot be easily encoded are problematised as 'noise' or 'error'. The subsequent categorisation of such 'errors' and the development of methods by which to reduce them has spawned entire fields of research and launched more than a few careers. However, the authors argue this approach is in fact the result of the asymmetric balance of power in favour of data practitioners, or perhaps more accurately, data users. As [Alain Desrosières](#) notes, 'statisticians justify their extensive efforts to whittle down any anomalies [arguing]: "our users would not tolerate our giving them inconsistent data"' (348).

It is therefore important for statisticians and policymakers alike to understand that non-responses or 'invalid' responses on the part of data subjects, far from being a purposeful reduction of data quality, are often the result of encountering statistical categories that do not reflect their lived reality (for example, the campaign to include a 'mixed race' category in the US census documented in [Mark One or More](#)). Even when data subjects do purposefully supply 'erroneous' responses, these are not driven by sheer perversity, but rather expressions of (non)acceptance of statistical categories that represent subjects' 'capacity to act and influence (or subvert) how they are categorized' (207).

Such actions are often taken as an attempt to escape what Ruppert and Scheel call the 'double edge of enumeration': 'being counted is simultaneously a precondition of recognition and in turn government support, but also makes possible intrusive and potentially harmful governing interventions such as eviction, detention, or deportation' (92). The recent debate in the UK over Clause 9 of the Nationality and Borders Bill has brought to light how [both these concerns exist for certain, often marginalised, groups](#). In states that enact the nation on the basis of inherited nationality, second and third generation individuals, despite holding birthright citizenship, are categorised officially as a type of 'other'. These individuals consider themselves full members of the nation, but may find that socially, or even legally, they are not. Understanding these concerns and accepting them as legitimate and thus worthy of consideration in the production of future statistics is necessary to protect individual rights as well as broader democratic institutions.

Data Practices's detailed case studies of the practice of statistical data production should be included on the syllabus of any introductory statistics, data science or research methods course. However, it is its thoughtful consideration of quantitative data production from the perspective of the data subject, beyond the (often) superficial consideration of data privacy and security, that will earn it pride of place on my bookshelf.

Note: This review gives the views of the author, and not the position of the LSE Review of Books blog, or of the London School of Economics and Political Science.

Banner Image Credit: [Pixabay CCO](#).
