



# A Primer on Powerful Numbers

Selected Readings in the  
Social Study of Public Data  
and Official Numbers

Dan Bouk,  
Kevin Ackermann,  
& danah boyd

There is a class of crucial, powerful numbers that exist in the modern world and upon which many societies rely. Examples of these official numbers, this public data, include:

- complete population counts as generated by state-run censuses;
- indices, metrics, and rates published by bureaucrats (such as cost-of-living indices, GDP, crime rates, unemployment rates, or infection and hospitalization rates) or by news organizations (as in the Dow Jones Industrial stock average or baseball batting averages);
- election tallies certified by state officials and legislatures;
- climate data released in reports by the Intergovernmental Panel on Climate Change (IPCC);
- ratings and scores produced by private organizations like Experian, or S&P, or the Educational Testing Service (ETS).

The thing that makes an official number different from any other number is that it is given with authority and always there for the taking.

The word “data” comes from the Latin for “the givens.” Whether one is building an argument or training a machine-learning algorithm, one starts with some set of givens: with data. Some data sets are hand-made or home-brewed. Official data sets come out of offices and are imbued with the authority of those in power. They are offered publicly with the expectation that when given, they will be taken. They are a shared basis for the building of arguments and algorithms.

Scholars in a host of fields—the history of quantification; economic sociology; science & technology studies, etc.—have studied the production of official numbers and developed a basic understanding of how they work in the world. This primer introduces readers to some of the work done by such scholars and highlights the following key arguments:

- ① **Modern Societies Are Built to Trust in Official Numbers. They Even Let Official Numbers Make Key Decisions.**
- ② **Official Numbers Are Made, Not Found.**
- ③ **When Things Are Going Well, We Forget That Official Numbers Had to Be Made.**
- ④ **Institutions Make Public Data and They Make Data Public.**
- ⑤ **Official Numbers Are Political.**
- ⑥ **Consensus on Official Numbers Requires Work. (It isn’t certain that the givens will be taken.)**

This primer is a syllabus organized around these six arguments. Like all syllabi, it is incomplete. It is parochial, limited to English-language scholarship, and

focused on Anglo-American historical cases: it is a place to start, but not the place to end. The following arguments do not stand alone, but rather are tightly intertwined and overlapping. Some points raised to discuss one argument could just as easily appear in another. The primer is intended to serve students or curious professionals who work with data, build algorithms (or critique them), or shape policy. It is for those who want to be able to read the marks that societies make on our numbers and see more clearly the ways that numbers shape our societies.

(We thank Jenna Burrell, Chris Phillips, and Ranjit Singh for offering comments and suggestions on this primer and Patrick Davison for his expert editing. The DAL Team at Data & Society offered early feedback and encouragement. This primer was produced with financial support from Alfred P. Sloan Foundation.)

---

# 1.

## Modern Societies Are Built to Trust in Official Numbers. They Even Let Official Numbers Make Key Decisions.

---

### GET STARTED READING:

Wendy Nelson Espeland and Mitchell L. Stevens, “A Sociology of Quantification,” *European Journal of Sociology* 49, no. 3 (2008): 401-436.

Theodore M. Porter, *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life* (Princeton, NJ: Princeton University Press, [1995] 2020).

Madeleine Clare Elish, “Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction,” *Engaging Science, Technology, and Society* 5 (2019): 40-60.

Ruha Benjamin, *Race After Technology: Abolitionist Tools for the New Jim Code* (New York: Polity, 2019).

Numbers bear special authority in the modern world. According to Espeland and Stevens, “The authority of numbers may be vested (1) in our sense of their accuracy or validity as representations of some part of the world; (2) in their usefulness in solving problems; (3) in how they accumulate and link users who have investments in the numbers; or (4) in their long and evolving association with rationality and objectivity” (417).

Official numbers retain authority so long as they make sense and jibe with other evidence; are useful within established or emerging contexts; and support and move within networks of users—and so long as rationality and objectivity remain cultural, economic, and political values.

Theodore Porter’s landmark book *Trust in Numbers* explains how and why official numbers have become so important over the last two centuries. First, he argues, it is because numbers are “technologies of distance” that serve (along with other communication technologies) to tie together widespread systems and bureaucracies. Second, Porter argues that numbers have increasingly been integrated into decision-making systems such that the numbers seem to make decisions automatically and seem to decrease our reliance on the judgment and discretion of officials or experts.

Letting numbers (appear to) make decisions—about who gets credit, who gets out of jail, who gets food stamps, who gets insurance, whose lands will be intentionally flooded, who works when, whose pay rises (or fails to)—insulates decision makers from political pressure or blow back. This produces what Madeleine Clare Elish calls “a moral crumple zone.” By translating political decisions into technical problems with

---

supposedly neutral, objective solutions, those in charge can exercise power without moving a finger. Make no mistake, though: making decisions with numbers does not eliminate judgment, politics, or bias. In fact, it often doubles down on existing inequities. For instance, Ruha Benjamin explains how seemingly objective technologies enact anti-Black racism, a “New Jim Code” that worked through both old-school statistics and the most modern of predictive algorithms. Official numbers encode centuries of racist policing practices or discriminatory risk models, which leads even the most sophisticated algorithm to drag the injustices of the past deep into our future.

### READ MORE:

(These readings are ordered to correspond, roughly, with the related topics of the text above.)

**Deborah Stone**, *Counting: How We Use Numbers to Decide What Matters* (New York: Liveright, 2020).

**Cathy O’Neil**, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (New York: Crown, 2016).

**Catherine D’Ignazio and Lauren F. Klein**, *Data Feminism* (Cambridge: MIT Press, 2020).

**Virginia Eubanks**, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (New York: St. Martin’s Press, 2018).

**Angèle Christin**, *Metrics at Work: Journalism and the Contested Meaning of Algorithms* (Princeton: Princeton University Press, 2020).

**Josh Lauer**, *Creditworthy: A History of Consumer Surveillance and Financial Identity in America* (New York: Columbia University Press, 2017).

**Martha Poon**, “From New Deal Institutions to Capital Markets: Commercial Consumer Risk Scores and the Making of Subprime Mortgage Finance,” *Accounting, Organizations and Society* 34 (2009): 654-674; and Martha Poon, “Scorecards as Devices for Consumer Credit: the Case of Fair, Isaac & Company Incorporated,” *Sociological Review* 55, supplement 2 (2007): 284-306.

**Bernard E. Harcourt**, *Against Prediction: Profiling, Policing, and Punishing in an Actuarial Age* (Chicago: University of Chicago Press, 2006).

**Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner**, “Machine Bias: There’s Software Used across the Country to Predict Future Criminals. And It’s Biased against Blacks,” *ProPublica* 23 May 2016.

**Cynthia Rudin, Caroline Wang, and Beau Coker**, “The Age of Secrecy and Unfairness in Recidivism Prediction,” *HDSR* 2.1 (2020).

**Nick Estes**, *Our History Is the Future: Standing Rock versus the Dakota Access Pipeline, and the Long Tradition of Indigenous Resistance* (New York: Verso, 2019).

---

**Dan Bouk**, *How Our Days Became Numbered: Risk and the Rise of the Statistical Individual* (Chicago: University of Chicago Press, 2015).

**Thomas A. Stapleford**, *The Cost of Living in America: A Political History of Economic Statistics, 1880-2000* (Cambridge: Cambridge University Press, 2009).

---

## 2. Official Numbers Are Made, Not Found.

---

### GET STARTED READING:

Lisa Gitelman, ed. *Raw Data Is an Oxymoron* (Cambridge, MA: MIT Press, 2013).

Jacqueline Wernimont, *Numbered Lives: Life and Death in Quantum Media* (Cambridge, MA: MIT Press, 2018).

Ian Hacking, “Biopower and the Avalanche of Printed Numbers,” *Humanities in Society* 5, no. 3 & 4 (1982): 279-294.

Benedict Anderson, “Census, Map, Museum,” in *Imagined Communities: Reflections on the Origin and Spread of Nationalism* (New York: Verso, 1991), 163-185.

Khalil Gibran Muhammad, *The Condemnation of Blackness: Race, Crime, and the Making of Modern Urban America* (Cambridge, MA: Harvard University Press, [2010] 2019).

According to a leading book in the social studies of data, “Raw Data Is an Oxymoron.” That’s because all data comes more or less cooked, more or less structured by a set of questions, assumptions, and procedures used in collecting and making it.

There is (sadly?) no “official number” tree from which ready-to-eat datasets can be plucked. Instead, we rely on agencies and offices to produce numbers. Some such operations are small—think of the tabulations made for a village election—these are the artisanal pickles of the official number world. In contrast, massive operations like state censuses or global climate figures result from industrial-scale processes—they are the factory-farmed sausages of official numbers. And, as is true in the case of actual pickles and sausages, it can be shocking for an outsider to see how numbers actually get *made*.

Scholars of quantification investigate how particular times and circumstances make numbers, make them important, and influence the form those numbers take. Media scholar Jacqueline Wernimont, for instance, has investigated the history of London’s “bills of mortality,” first published in 1603 and made significant by the dangers of plague in the mid-17th century. Wernimont’s analysis shows how a fundamentally chaotic problem (plague) inspired early modern clerks and men of science (working in guilds or societies with royal sanction) to build tabular representations of mortality in ways that conveyed a calming sense of order and mastery.

It’s easy to forget that statistical tables are technologies, and that they have a history. (Graphs do too! Drawings or prints on paper may not feel like technologies, but they are actually sophisticated tools for displaying and analyzing data—not so different from data

---

visualizations displayed on an LCD screen today.) It's even easier to forget that those tables don't just convey information, they convey a sense of stability and certainty—they do emotional work. Numbers and their authors too often fool us into thinking that they have nothing to do with feelings, and they can fool us in other ways too. Tables tend to disguise some of their limitations and their origins. As Wernimont explains, the plague-era bills of mortality erased the labor of women (who did much work gathering data on individual bodies) and the suffering of people of color (since they were people not to be counted or included in the tables). Statistical tables and even the concept of statistical regularities explained by statistical laws grew in public prominence because of the mid-17th century plague. Today, the COVID pandemic—which has been made manifest by a flood of numbers, charts, and simulations—has likely changed the landscape of official numbers in ways that only close study will reveal.

In the intervening centuries between the plague years and the COVID years, factory-farmed data arrived on the scene. The philosopher Ian Hacking has identified an “avalanche of printed numbers” that took place in the early nineteenth century marking the beginning of the transition across Europe toward mass-produced statistics. According to Hacking, state bureaucracies and private charities alike produced data as part of a program to better understand the poor and oppressed—in the hopes that such understanding would lead to improved conditions and behavior, and stay any revolutionary impulses of the disgruntled masses in an industrializing society. This avalanche did not just lead to massive piles of data—it made categories for thinking about societies that shaped how people saw themselves and one another. For those who claim the new era of big data is threatening human nature, perhaps it will come as a relief that the making of data has been remaking what it means to be human for hundreds of years.

When we accept that data are made, not found, it becomes necessary to think about the systems that make them. Much of the history of big, official data belongs to a larger history of expanding, imperial nation states and booming corporations (led by railroads in the nineteenth century, then industrial multinationals, and today by Silicon Valley behemoths). Benedict Anderson, writing about colonial governance, explains the similar functions that governments' censuses, maps, and museums all served. He notes a shared commitment to “a totalizing classificatory grid, which could be applied with endless flexibility to anything under the state's real or contemplated control: peoples, regions, religions, languages, products, monuments and so forth...[that grid] was bounded, determinate, and therefore—in principle—countable.” The goal of control and the principle of complete, comprehensive countability have driven the production of official numbers for a long time.

Data might be “the givens,” but facts, which come from the latin *facere*, are “that which has been made.” Statistics take a set of givens and transmute their embedded social, economic, and political hierarchies into objective facts. Thus, for instance, economists and statisticians made criminality into a Black “race trait” at the turn



---

of the twentieth century, according to research by Khalil Gibran Muhammad. Racialized public crime data then, in the early twentieth century, became the basis for predictions of criminality that have shaped policing, punishment, and parole for decades. Those systems, now more complex, but still supposedly objective, remain in operation. Historical scholarship has helped draw attention to the people, politics, and prejudices that shaped both data and fact.

Even if numbers appear to come in an avalanche, they are—in the end—made. Official numbers are not the product of natural forces. They are made by entities with power and they play powerful roles in society. Anyone committed to making equitable data, better facts, and reliable numbers must look closely at the social processes and influential actors behind the numbers.

## READ MORE:

**Paul N. Edwards**, *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming* (Cambridge, MA: MIT Press, 2010).

**Lorraine Daston**, “Super-Vision: Weather Watching and Table Reading in the Early Modern Royal Society and Académie Royale des Sciences,” *Huntington Library Quarterly* 78, no. 2 (2015): 187-215.

**Carla Bittel, Elaine Leong, and Christine von Oertzen**, editors, *Working with Paper: Gendered Practices in the History of Knowledge* (Pittsburgh: University of Pittsburgh Press, 2019).

**Thomas L. Hankins**, “A ‘Large and Graceful Sinuosity’: John Herschel’s Graphical Method,” *Isis* 97, no. 4 (Dec. 2006): 605-633.

**Judy L. Klein**, *Statistical Visions in Time: A History of Time Series Analysis, 1662-1938* (New York: Cambridge University Press, 1997).

**Whitney Battle-Baptiste and Britt Rusert**, eds., *W. E. B. Du Bois’s Data Portraits: Visualizing Black America* (New York: Princeton Architectural Press, 2018).

**Kirsten Ostherr**, “How Do We See COVID-19? Visual Iconographies of Racial Contagion,” *American Literature* 92, no. 4 (2020): 707-722.

**Jacqueline Wernimont**, “Media Infrastructure Is and Has Always Been a Matter of Life and Death,” *items* 15 April 2021.

**Christopher J. Phillips**, “The Curve,” *Historical Studies in the Natural Sciences* 51, no. 1 (2021): 159-161.

**Dan Bouk**, “Materializing COVID,” *Isis* 111, no. 4 (2020): 783-786.

**Richard White**, *Railroaded: The Transcontinentals and the Making of Modern America* (New York: Norton, 2011).

**Timothy Mitchell**, *Rule of Experts: Egypt, Techno-Politics, Modernity* (Berkeley: University

---

of California Press, 2002).

Ian Hacking, “Making Up People” *London Review of Books* 28, no. 16 (2006): 23–26.

Caitlin Rosenthal, “Lessons from the ‘Data Exhaust’ of Plantation Slavery,” *Harvard Data Science Review* 3, no. 2 (2021).

Eli Cook, *The Pricing of Progress: Economic Indicators and the Capitalization of American Life* (Cambridge, MA: Harvard University Press, 2017).

Mary Poovey, *A History of the Modern Fact: Problems of Knowledge in the Sciences of Wealth and Society* (Chicago: University of Chicago Press, 1998).

Simone Browne, *Dark Matters: On the Surveillance of Blackness* (Durham: Duke University Press, 2015).

---

## 3.

# When Things Are Going Well, We Forget That Official Numbers Had to Be Made.

---

### GET STARTED READING:

Christopher J. Phillips, *Scouting and Scoring: How We Know What We Know about Baseball* (Princeton: Princeton University Press, 2019).

Caitlin Zaloom, “Ambiguous Numbers: Trading Technologies and Interpretation in Financial Markets,” *American Ethnologist* 30, no. 2 (2003): 258-272.

Geoffrey C. Bowker and Susan Leigh Star, *Sorting Things Out: Classification and Its Consequences* (Cambridge, MA: MIT Press, 1999).

After official numbers are made, they begin to move, shedding the context that created them. The circumstances of their birth grow hazy or are forgotten entirely. The users of the data suspend the statistical disbelief that comes from seeing the sausage get made and they get to calculating or arguing, as if the numbers were simply true—or true enough, true by agreement. Soon enough, they ascribe new meanings to the numbers before them or inject those numbers into new contexts.

Imagine this: two sports fans debate the merits of their favorite players, setting the accomplishments of one against the other. More often than not, those cases will rest on a comparison of one player’s statistical record against the other’s. Sports stats would seem to be among the simpler sort of official numbers: just count the baseball batter’s hits and at-bats, etc. How hard is that? Yet in his careful study of the production of baseball data, Chris Phillips reveals a surprisingly rickety system for gathering official statistics—one often criticized from all sides, one dependent on official scorers working with little standardization employed under various ad hoc arrangements, whose decisions frequently draw criticism. Yet Phillips also reveals that it has mostly worked: “Official scorers have consistently managed to produce numbers through contestable judgments that, remarkably, have come to be treated as authoritative and objective statistics” (64). When it comes time to have their fight, those sports fans won’t worry about the subjectivity subsumed in the statistics they toss at one another.

The users of official numbers add subjective meanings to those numbers as they use them, and not just while arguing about sports. Anthropologist Caitlin Zaloom went into the digital “pits” to observe the methods that London traders used as they speculated

---

on futures contracts, betting millions upon millions of dollars with each keystroke. Each trader had a terminal that scrolled feeds of shifting numbers, representing the market. But it struck Zaloom that these traders transformed the numbers they worked with. “Traders who use financial technologies do not take up numbers as objective descriptions of supply and demand,” she wrote. Instead, they “seek out nonquantitative information that is located within the market numbers” (269). They interpret the flow of numbers in new ways, adding new layers of meaning to the stream of digits. They see in the numbers a series of social interactions, and they look for their advantage, for the right moment to get in and make a killing.

The uses of official numbers in baseball and big finance both illustrate that official numbers result from infrastructures, from interconnected technical systems that bind users of the numbers to complex networks of observers and complicated processes for turning observations into numbers. Ethnographers Star and Ruhleder defined “infrastructure” in terms of its far-reaching “embeddedness” with communities of users, where infrastructure embodies standards and follows or enables conventions that are meaningful to those communities. Official numbers come to us—whether in avalanches or as the new oil—through data infrastructures.

Users of official numbers—whether they’re sports fans or futures traders, or economists looking at GDP, or epidemiologists looking at hospitalization rates, or risk managers looking at bond ratings—learn how to use and interpret those numbers, and if the infrastructure is working as usual, they don’t have to trouble themselves with the nitty-gritty details of how those numbers get made. Sometimes the infrastructure includes reminders of the uncertainty baked into the numbers, like error bars or uncertainty intervals, measures of statistical significance, and explanations of “total survey error,” but these too serve to make it so that the numbers’ users don’t need to think too much about the specifics of how the numbers were made. Such is the nature of infrastructures: they are usually invisible (“transparent”) to users, according to Star and Ruhleder. They only become “visible upon breakdown,” and are then repaired in piecemeal, “modular” fashion. No one notices an infrastructure until it breaks, and no one thinks about where official numbers come from until a controversy kindles.

Pollsters in the United States in 2016 (or 2020) made their phone calls, asked their questions, and begged for responses. High profile analysts aggregated their results, built data models from them, and generated predictions about who would end up winning the presidency. The numbers generated through polling and prediction moved far and wide. The processes that made them drew much less attention. That changed the moment that election results differed substantially from the polls and prediction models. The polls and data models trundle along, until they are proven spectacularly wrong.

---

## READ MORE:

- Martha Lampland and Susan Leigh Star, *Standards and Their Stories: How Quantifying, Classifying, and Formalizing Practices Shape Everyday Life* (Ithaca: Cornell University Press, 2009).
- Ranjit Singh, “Study the Imbrication: A Methodological Maxim to Follow the Multiple Lives of Data,” in *Lives of Data: Essays on Computational Culture in India* edited by Sandeep Mertia (Amsterdam: Institute of Network Cultures, 2020): 51-59.
- Alain Desrosières, “How Real Are Statistics? Four Possible Attitudes,” *Social Research* 68, no. 2 (2001): 339–355.
- David Ribes and Steven J. Jackson, “Data Bite Man: The Work of Sustaining a Long-Term Study,” in *Raw Data Is an Oxymoron* edited by Lisa Gitelman (Cambridge, MA: MIT Press, 2013), 147-166.
- Robert Groves and Lars Lyberg, “Total Survey Error: Past, Present, and Future,” *Public Opinion Quarterly* 74, no. 5 (2010): 849-87.
- Theodora Dryer, “Designing Certainty: The Rise of Algorithmic Computing in an Age of Anxiety 1920 – 1970,” (PhD Dissertation, UC San Diego, 2019).
- Phaedra Daipha, *Masters of Uncertainty: Weather Forecasters and the Quest for Ground Truth* (Chicago: University of Chicago Press, 2015).
- Jamie L. Pietruska, *Looking Forward: Prediction and Uncertainty in Modern America* (Chicago: University of Chicago Press, 2017).

---

## 4.

# Institutions Make Public Data and They Make Data Public.

---

### GET STARTED READING:

Caitlin Rosenthal, *Accounting for Slavery: Masters and Management* (Cambridge, MA: Harvard University Press, 2018).

William Deringer, *Calculated Values: Finance, Politics, and the Quantitative Age* (Cambridge, MA: Harvard University Press, 2018).

Emily Klancher Merchant, *Building the Population Bomb* (New York: Oxford University Press, 2021).

Joanna Radin, “Digital Natives’: How Medical and Indigenous Histories Matter for Big Data,” *Osiris* (2017): 43-64.

Any institution that has an office—or something like an office—can produce nominally “official” numbers. But not all numbers generated in offices become public numbers. Only numbers that get out into the public can attain that status. And the history of numbers and data reminds us that institutions have to choose to make their numbers public, to make sure they are there for the taking.

Some of the earliest and most important kinds of public data were originally produced by early modern businesses addressing themselves to their trade partners, to their investors, to regulatory governments, and even to God. Merchants engaged in long-distance maritime trade in the thirteenth century were legally required by major ports like Venice and Barcelona to employ scribes to create official records of voyages and inventory. From such laws and mercantile practices emerged formal accounting systems, which shifted from ship to shop over the following centuries and eventually took root, as Caitlin Rosenthal has explained, in the most complicated and elaborate operations of the early modern Atlantic world: slave-exploiting, sugar-producing plantations. When the merchants and planters of the slave-economy opened their ledger books to investors and regulators, they effectively transformed their private economic numbers into powerful public records.

European states moved from mostly encouraging the production of official numbers by private individuals toward producing their own in the early modern period—imperial Chinese states already had long-established traditions of data gathering by that point. Sometimes, as in France under Louis XIV’s administrator Jean-Baptiste Colbert, such data were generated to serve the centralized state, and the state alone. In other cases, opposition

---

parties or those seeking to access the state's secrets produced new, more public-facing numbers. Such was the situation, explains historian William Deringer, that gave birth to “Political Arithmetick” and calculation as a powerful political tool in England at the dawn of the eighteenth century. With the emergence of the modern nation-state in Europe, the production of data surged further and that data became more frequently public. The secrets of the centralized state became the official numbers that we now call “statistics.”

As the nineteenth turned to the twentieth century, offices (of states, of corporations, of labor unions, or farmers' cooperatives) grew in size and scope. Their struggle with one another for power generated many more official numbers. As corporations merged to create multidivisional, multinational firms, states developed new techniques for the managing of markets and the well-being of workers, citizens, and subjects. Businesses generated data about their operations for regulators, stock markets, and economic censuses—data that made possible new kinds of forecasting and new concepts like the “business cycle.” Such data could also be used to limit fraud and speculation. Economists generated new economic indices establishing average “standards of living” or the size of “the Economy.” Pollsters and market researchers generated data on the desires and opinions of consumers and citizens. Private companies, the International Labor Organization, and academic researchers all contributed to measures of the inputs and outputs necessary to maintain workers' bodies.

Official numbers increasingly tackled global scales with the development of global governance. After World War II, new international governance structures and institutions (the United Nations, the Bretton Woods system, the superpowers and their satellites) stimulated a global standardization and expansion of statistical reporting. Emily Merchant has demonstrated, for instance, the role of the United Nations in establishing demographic standards that, in turn, generated the first reliable (and to some alarming) global population totals. The proliferation of population data and with that data, simulations preaching the economic benefit of population restriction, inspired aggressive birth control in many post-colonial states. Similarly, global climate data relied on international cooperation and Cold War surveillance systems, interpreted with powerful computer models, to create official figures, which proved to be both alarming and controversial.

The invention of the electronic computer and an increased reliance on statistical techniques aided by computers (random sampling, multivariate regression, data modeling, simulation) also stimulated the production of and demand for finer grained data, often geolocated or tied to individuals over time. With the rising power of financial institutions toward the end of the century, both firms and people came to be seen more as assets than as distinct, productive entities. Digitized official data, like for instance, financial data, might become almost ephemeral—flashing across a trader's screen just long enough to strike a deal. Other public data might,

---

thanks to widespread privatization, become differently ephemeral, as states turned over some numbers directly to private interests, leaving the public in the dark.

While official numbers are made public by institutions, they often represent people. They are, in a very real sense, personal. Data given for one purpose can, after that data has been published and made official, be turned to new purposes without the consent or even consultation of those represented in the numbers. Joanna Radin's path-breaking article, "Digital Natives," shows how a dataset gathered ostensibly to improve the health of Indigenous people became a generic set of numbers used to train machine learning algorithms. She shows how colonial power structures transformed personal data into public—even ubiquitous—numbers. It would be better, she argues, if communities who had offered data had a seat at the table and real power to decide how to make the data work for them and for those around them.

## READ MORE:

**Bruce G. Carruthers and Wendy Nelson Espeland**, "Accounting for Rationality: Double-Entry Bookkeeping and the Rhetoric of Economic Rationality," *American Journal of Sociology* 97, no. 1 (1991): 31-69.

**Katherine McKittrick**, "Mathematics Black Life," *Black Scholar* 44, no. 2 (2014): 16-28.

**Lorraine Daston**, *Classical Probability in the Enlightenment* (Princeton, NJ, 1988).

**Jonathan Levy**, *Freaks of Fortune: The Emerging World of Capitalism and Risk in America* (Cambridge, MA: Harvard University Press, 2012).

**Timothy Shenk**, "Inventing the American Economy," (Ph.D. diss, Columbia University, 2016).

**Walter A. Friedman**, *Fortune Tellers: The Story of America's First Economic Forecasters* (Princeton, NJ: Princeton University Press, 2014).

**Daniel Abramson Hirschman**, "Inventing the Economy: Or: How We Learned to Stop Worrying and Love the GDP" (PhD diss., University of Michigan, 2016).

**Sarah E. Igo**, *The Averaged American: Surveys, Citizens, and the Making of a Mass Public* (Cambridge, MA: Harvard University Press, 2008).

**Sarah E. Igo**, *Known Citizen: A History of Privacy in Modern America* (Cambridge, MA: Harvard University Press, 2018).

**Dana Simmons**, *Vital Minimum: Need, Science, and Politics in Modern France* (Chicago: University of Chicago Press, 2015).

**Nick Cullather**, "The Foreign Policy of the Calorie," *American Historical Review* 112, no. 2 (2007) 337-364.

**Theodore M. Porter**, *Genetics in the Madhouse: The Unknown History of Human Heredity*



---

(Princeton: Princeton University Press, 2018).

**Coreen McGuire**, *Measuring Difference, Numbering Normal: Setting the Standard for Disability in the Interwar Period* (Manchester: Manchester University Press, 2020).

**Michelle Murphy**, *The Economization of Life* (Durham: Duke University Press, 2017).

**Jill Lepore**, *If Then: How the Simulmatics Corporation Invented the Future* (New York: Liveright, 2020).

**Paul Edwards**, *The Closed World: Computers and the Politics of Discourse in Cold War America* (Cambridge: MIT Press, 1996).

**Caley Horan**, *Insurance Era: Risk, Governance, and the Privatization of Security in Postwar America* (Chicago: University of Chicago Press, 2021).

**Claire Fontaine and Kinjal Dave**, *Spectrum of Trust in Data: New York City Parents Navigating School Choice* (Data & Society, 2018).

**Kadija Ferryman**, “Reframing Data as a Gift,” SSRN (April 21, 2017).

---

## 5. Official Numbers Are Political.

---

### GET STARTED READING:

Emmanuel Didier, *America by the Numbers: Quantification, Democracy, and the Birth of National Statistics* Translated by Priya Vari Sen (Cambridge, MA: MIT Press, 2020).

Michelle Murphy, *Sick Building Syndrome and the Problem of Uncertainty: Environmental Politics, Technoscience, and Women Workers* (Durham, NC: Duke University Press, 2006).

Arunabh Ghosh, *Making it Count: Statistics and Statecraft in the Early People's Republic of China* (Princeton, NJ: Princeton University Press, 2020).

Official numbers are always political, even when they are not turned to partisan ends. They are political because they enable or encourage some kinds of debate or some distributions of resources, while preventing or discouraging others. Using official numbers, political combatants can make their arguments, stake their claims, push their agendas, or enforce their decisions.

Official numbers can be political at the macro level, where certain kinds of numbers bring aggregates into being that can then be measured, managed, and manipulated. In Emmanuel Didier's *America by the Numbers*, it's evident that the rise of the welfare state brought into being a distinct kind of statistical America—one that required and made national measures of wheat production and measures of unemployment rates. Statisticians produced official numbers that captured America “as a whole,” an object to be fine-tuned by interventionist technocrats.

The politics of official numbers operate even more frequently at the micro level, as a means for listing (and often valuing) all ways of existing that are acceptable for prevailing systems of commerce, investment, and governance. Official numbers, in such cases, affix to each person or product their proper rank or place. Intelligence testing is the classic example of this mode of politics. While IQ was intended as a kind of group measure for building aggregates, it in practice more often serves to sort populations, slicing and dicing them, placing each individual into some narrow spot (their score) which will open some doors and close others.

There are politics too, to be sure, in silences. Some statistics never get prepared, or are prepared and then aren't published. A liberal politics that values privacy—or that contains interest

---

groups that fight for privacy—may prevent a census office from releasing personal information. An embarrassed government could, for rather different reasons, refuse to allow certain kinds of numbers to be gathered, or suppress them after they have been prepared. Examples include the 1937 Soviet census that was completed and then so thoroughly buried that its architects were executed. Private interests, perhaps especially corporate interests, also possess both a desire and capacity to create silences. Michelle Murphy demonstrated how both state and private actors constructed a system for detecting chemical exposure that made low-level exposures invisible. “When ignorance has been generated not just accidentally but also purposively,” she wrote, “...then the struggle of ordinary people to understand their bodies and the consequential, sometimes deliberate, undermining of their effort resonates with a political, and not just poignant, valence.” Murphy’s work, and the work of other scholars, seeks to understand both the making of silences and the work of others, often of those most harmed by silence, to fill the map and cope as best they can.

Of course, official numbers are sometimes political in a partisan sense—they can be directly linked to a particular party in power. In his book, *Making It Count*, Arunabh Ghosh explained how the People’s Republic of China halted the seemingly inexorable rise of probability as a tool for making sense of state data in the 1950s. Statistics produced by surveying and sampling became “bourgeois” statistics and had to be set aside. This sort of thing has happened all over the world, and especially with censuses. In the United States in the 1990s, to take another prominent instance, probabilistic methods became subject to partisan contention quashing an effort to statistically adjust the counts used in allocating political power. There doesn’t have to be anything inherent in any set of official numbers for them to become partisan, to be made partisan.

## READ MORE:

**Alain Desrosières**, *The Politics of Large Numbers: A History of Statistical Reasoning* (Cambridge, MA, 1998).

**William Alonso and Paul Starr**, editors, *The Politics of Numbers* (New York: Russell Sage Foundation, 1987).

**Colin Koopman**, *How We Became Our Data: A Genealogy of the Informational Person* (Chicago: University of Chicago Press, 2019).

**Alice O’Connor**, *Poverty Knowledge: Social Science, Social Policy, and the Poor in Twentieth-Century U.S. History* (Princeton, NJ: Princeton University Press, 2001).

**Aryn Martin and Michael Lynch**, “Counting Things and Counting People: The Practices and Politics of Counting,” *Social Problems* 56, no. 2 (2009): 243-266.

**Erik Linstrum**, *Ruling Minds: Psychology in the British Empire* (Cambridge, MA: Harvard University Press, 2016).

- 
- John Carson**, *The Measure of Merit: Talents, Intelligence, and Inequality in the French and American Republics, 1750-1940* (Princeton, NJ: Princeton University Press, 2006).
- Catherine Merridale**, “The 1937 Census and the Limits of Stalinist Rule,” *Historical Journal* 39, no. 1 (1996): 225-240.
- S.A. Aluko**, “How Many Nigerians? An Analysis of Nigeria’s Census Problems, 1901-63,” *Journal of Modern African Studies* 3, no. 3 (1965): 371-392.
- Ulrich Beck and Peter Wehling**, “The Politics of Non-Knowing: An Emerging Area,” in *The Politics of Knowledge* (Routledge: 2011).
- Martha Lampland**, “False numbers as formalizing practices,” *Social Studies of Science* 40, no. 3 (2010): 377-404.
- Molly Farrell**, *Counting Bodies: Population in Colonial American Writing* (New York: Oxford University Press, 2016).
- Margo J. Anderson and Stephen E. Fienberg**, *Who Counts?: The Politics of Census-Taking in Contemporary America* (New York: Russell Sage Foundation, 1999).
- Martha Hodes**, “Fractions and Fictions in the United States Census of 1890,” in *Haunted by Empire* edited by Ann Laura Stoler, (Durham: Duke University Press, 2006), 240-270.
- Margo J. Anderson**, *The American Census: A Social History* (New Haven, CT: Yale University Press, 2015) Second revised edition.

---

## 6.

# Consensus on Official Numbers Requires Work. (It isn't certain that the givens will be taken.)

---

### GET STARTED READING:

Paul Starr, “The Sociology of Official Statistics,” in *The Politics of Numbers* edited by William Alonso and Paul Starr (New York: Russell Sage Foundation, 1987).

Thomas Kuhn, *The Structure of Scientific Revolutions* (Chicago: University of Chicago Press, 1962).

William James, *Pragmatism: A New Name for Some Old Ways of Thinking* (New York: Longmans, Green, and Co., 1907).

Naomi Oreskes and Erik M. Conway, *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming* (New York: Bloomsbury Press, 2010).

Just because an official number makes its way in the world, doesn't mean that number was necessarily accurate, precise, or right. It means the number is official. “The axiom of modern computer research—garbage in, garbage out—does not exactly apply when statistics are official,” speculated Paul Starr in 1987. “In that case the rule may be *garbage in, consensus out.*” When a commodity can be used to pay taxes and enjoys near universal fungibility, we call it cash or money. When numbers can be plugged into spreadsheets, trotted out in arguments, and employed by bureaucrats with near universal acceptance, we call them official numbers. Official numbers are official numbers if, when they're given, they're *taken*—by other officials and plenty of other people or processes too.

This approach to official numbers builds on a long tradition in both pragmatic philosophy and in STS research demonstrating that purely technical factors cannot explain (on their own) how or why one scientific theory or practice wins out or why a scientific controversy ends. In the 1962 *Structure of Scientific Revolutions*, Thomas Kuhn argued that it is hard to prove theories wrong, no matter what new challenges are posed to them or new evidence discovered. So how is it that any theory gets undone? Kuhn argued a theory can be displaced when an alternative both fits the facts well and gives researchers in the field, laboratory, and library a new set of productive questions to pursue. Later researchers looked at the most famous experiments in the most famous “scientific revolution” (in early modern England) and argued that the success of the revolutionaries could not simply be attributed to the fact that they were right. Their rightness was the result of a historical drama, not its cause. More recently, in work on “performativity” scholars showed that some official economic numbers generated the market regularities they

---

purported to describe.

A deep strain of philosophical pragmatism informs this tradition. In the words of William James, “Theories thus become instruments, not answers to enigmas, in which we can rest.” Or “any idea upon which we can ride, so to speak; any idea that will carry us prosperously from any one part of our experience to any other part, linking things satisfactorily, working securely, simplifying, saving labor; is true for just so much, true in so far forth, true instrumentally.” Theories are true when they work. Instruments that hold water are sound. And, when applied to official numbers, the final proof of the data is in its use.

So are accuracy, precision, and quality beside the point? By no means. Those qualities are simply insufficient to assure any dataset becomes an official number.

This does not come as news to those with an interest in undermining consensus on important public issues, those who strategically sow doubt and mass produce ignorance. They recognize all too well that building a consensus takes work, and that even an existing consensus can be undermined. Using techniques developed by Big Tobacco, entities with financial and ideological interests in deregulation and the dismantling of official structures keeping industries honest have turned to dealing in doubt. They seek to impugn the quality of public numbers, or impugn the data makers, or to generate alternate numbers or “alternative facts”—hoping to prevent the numbers produced by government offices or scientists from becoming universally accepted. They want to prevent the fruits of sound science from acting as official numbers must, as givens that will be taken as true and used by policy makers.

This could be discouraging for those who work diligently to generate good data, to test the quality of their results, and to report them with suitable precision and an openness about error and uncertainty. It could be discouraging, or it could be simply a reminder: the job doesn’t end with the production of numbers. Official numbers only become official numbers when those who make them have done what is necessary to ensure their numbers get used, and used widely.

## READ MORE:

**Evelyn S. Ruppert**, “‘I Is; Therefore I Am’: The Census as Practice of Double Identification,” *Sociological Research Online* 13, no. 4 (2008).

**Wendy Nelson Espeland and Michael Sauder**, “Rankings and Reactivity: How Public Measures Recreate Social Worlds,” *American Journal of Sociology* 113, no. 1 (2007): 1–40.

**Steven Shapin and Simon Schaffer**, *Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life* (Princeton: Princeton University Press, 1985).

**Donald MacKenzie and Yuval Millo**, “Constructing a Market, Performing Theory: the

---

Historical Sociology of a Financial Derivatives Exchange,” *American Journal of Sociology* 109:1(July, 2003): 107-145.

**Donald MacKenzie**, *Inventing Accuracy: A Historical Sociology of Nuclear Missile Guidance* (Cambridge: MIT Press, 1990).

*Merchants of Doubt* (2014, feature film) dir. by Robert Kenner.

**Kate Brown**, *Manual for Survival: An Environmental History of the Chernobyl Disaster* (New York: W.W. Norton & Co., 2019).

**Robert N. Proctor and Londa Schiebinger (eds.)**, *Agnotology: The Making and Unmaking of Ignorance*, Stanford, CA (Stanford University Press, 2008).

Data & Society is an independent nonprofit research institute that advances new frames for understanding the implications of data-centric and automated technology. We conduct research and build the field of actors to ensure that knowledge guides debate, decision-making, and technical choices.

[datasociety.net](https://datasociety.net)  
[@datasociety](https://twitter.com/datasociety)

