

The History and Political Economy of Personal Data over the Last Two Centuries in Three Acts

by *Dan Bouk**

ABSTRACT

This essay uses the sociological concept of the “data double” to investigate the changing political economy of personal data in the United States in the nineteenth and twentieth centuries and up to today. It reviews secondary sources and examines select primary sources to argue for the ascendance of the double over the last two centuries and establishes a three-part periodization for personal data. In the first act, personal data worked mainly in aggregated form to help construct social imaginaries. From World War I until the 1970s, the power of personal data shifted toward the use of aggregates to fit individuals into the mass. Today, in the third act, individuals generate many data doubles that are commodified, capitalized, collected, celebrated, and often out of the control of those they represent.

Personal data is neither new nor newly powerful. The essays that precede this one have already proven as much. As Markus Friedrich shows, genealogists in early modern Europe worked feverishly to generate enough personal data to satisfy the demands of individuals and families who often needed that data to win or maintain political or economic status and position. In the modern era, personal data became even more crucial to the operation of states or companies, and to universities, too. Rebecca Lemov’s essay invites us to recognize the embeddedness of personal data in our institutions and practices and to reconsider Big Data as a profoundly social phenomenon through a close investigation of the datafied life of Don Talayesva, perhaps the mid-twentieth century’s most documented person. Joanna Radin’s essay, in turn, further evokes an embeddedness so deep that even in standardized machine learning data one can discover the lingering shadow of colonial exploitation. The power of personal data (to predict, among other things, manhole fires) derives in no small part,

* Department of History, Colgate University, 13 Oak Drive, Hamilton, NY 13346; dbouk@colgate.edu.

I thank the Max Planck Institute for the History of Science’s “Historicizing Big Data” working group and “History of Data” reading group, especially Lorraine Daston, Christine von Oertzen, Elena Aronova, David Sepkoski, Joanna Radin, Judith Kaplan, Markus Friedrich, and Etienne Benson. Two anonymous reviewers, Sarah Igo, Dan Hirschman, Danielle Keats Citron, and especially Martha Poon also provided helpful comments.

© 2017 by The History of Science Society. All rights reserved. 0369-7827/11/2017-0005\$10.00

her essay reminds us, from the now nearly invisible labor of those whose lives generated that data.¹

But while personal data is neither new nor newly powerful, its political economy has been anything but static. In fact, we live today amidst a series of profound changes in the political economy of personal data, as the preceding essays have also made clear. This essay aims to help us understand those changes better by telling a story in broad strokes of how personal data has worked in the world, and especially how it has worked within changing modes of power and regimes of valuation. I have focused my narrative on the last two hundred years and mostly on the United States, where my primary expertise lies. But the framework developed in this essay will be useful in other contexts and the narrative sweep familiar in other places.

To begin, I must offer a few general words on personal data. All personal data exists as some sort of inscription. In making personal data, someone writes something down or enters values into a spreadsheet or sketches lines on a chart. Each such inscription represents a person or multiple people, or even a prior such inscription. As we delve into the history and political economy of personal data, it is convenient to think of personal data inscriptions in two broad types or classes. First, we can think about those inscriptions that describe only a single individual. I find the concept of the “data double,” invented by the sociologists Kevin D. Haggerty and Richard V. Ericson, to be particularly compelling when thinking about such inscriptions.² Data doubles stand in for us in bureaucracies. They represent us on paper or in computer systems. They can be powerful, when, for instance, they convince a bank to provide us a mortgage or serve as evidence for a prosecution. Sociologists have offered other useful terms, too, like the concept of the “shadow body,” employed by Ellen Balka and Susan Leigh Star to think through the way hospitals create records about us that linger.³ Still others use terms like the “digital dossier,” “data exhaust,” or “data shadows,” but in this essay I will use data double exclusively.⁴

The second type of personal data inscription derives from the aggregation and analysis of data doubles. Such inscriptions can include charts, tables, maps, or even algorithms, and so there is no single obvious term we can use to lump them all together. Let’s call them data aggregates here.

Which have more power: data doubles or data aggregates? It’s tempting to side with data doubles. They relate directly to an individual and so link a person to whoever possesses that double, whether a government, corporation, or some other bureaucracy. When a student takes a standardized test, for instance, we think of the test and the raw score it yields as being very powerful, since the raw score represents (in a

¹ Markus Friedrich, “Genealogy as Archive-Driven Research Enterprise in Early Modern Europe”; Rebecca Lemov, “Anthropology’s Most Documented Man, Ca. 1947: A Prefiguration of Big Data from the Big Social Science Era”; Joanna Radin, “‘Digital Natives’: How Medical and Indigenous Histories Matter for Big Data,” all in this volume.

² Kevin D. Haggerty and Richard V. Ericson, “The Surveillant Assemblage,” *Brit. J. Sociol.* 51 (2000): 605–22. I use “inscription” in the spirit of Bruno Latour, *Science in Action: How to Follow Scientists and Engineers through Society* (Cambridge, Mass., 1987), chap. 6.

³ Ellen Balka and Susan Leigh Star, “Mapping the Body across Diverse Information Systems: Shadow Bodies and How They Make Us Human,” paper presented at the 4S Annual Meeting, Cleveland, Ohio, 2 November 2011, http://www.4sonline.org/files/print_program_abstracts_111007.pdf (accessed 8 February 2017).

⁴ See, e.g., Daniel J. Solove, *The Digital Person: Technology and Privacy in the Information Age* (New York, 2004).

DATA
DOUBLES
vs.
DATA
AGGREGATES
(individual vs.
population,
singular vs.
mass)

heavily limited and reduced fashion) the test taker. Yet later aggregations can be quite powerful as well. The average score for this student's school might well determine whether that school can remain open or how well it will be funded and thereby affect the school, its community, and the student's life through her school. To add a further complicating factor, the student's raw score may matter less than a percentile score derived by interpreting her performance in the context of all other students' performances—a new data double, in other words, may have to be constructed with reference to a data aggregate. There are no simple answers to questions of power, and so of economics and politics, relating to personal data.

Still, we can discern shifts over time in the distribution of power among data doubles and data aggregates and in the relationship of doubles to aggregates. First and foundationally, we can say that over the last two centuries, data doubles have become longer lived and more powerful than they previously were. With that trend in mind, we can further trace out three periods, three “acts” in the history of personal data, each characterized to a significant degree by the particular relationship of doubles to aggregates. The first act, which spanned the nineteenth century, saw the reign of aggregates derived from doubles and the invention of common means for ordering society. In the second act, which lasted until the 1970s, individuals were increasingly viewed through a statistical lens, which meant aggregates spawned a growing proportion of doubles used to fit people into their proper places in a mass society. Finally, in our own third act of personal data, more doubles met more (newer, nimbler) aggregates, inspiring some to dream of a dawning, fairer era of unbounded individualism. Yet such dreams required ignoring the persistent power of traditional aggregates in data-driven technologies—data, for instance, has not unmade race or racism. Nor has it unmade personal data's managerial drive or capitalist character. On the contrary, third-act dreams have fueled a more extensive extraction of value and a more intense (perhaps tragic) management of the people whom they were meant to free from their fetters.

ACT I: STRUCTURING THE SOCIAL

The story of personal data in the nineteenth century has been told before, but never as a story about personal data. First, it appeared as part of the history of statistics, occasioned to some extent by the desire of historians in the 1970s and 1980s (around the time the third act opened) to uncover histories that might shed light on the rise of mathematical statistics across the sciences and in independent academic programs beginning in the decades after World War II.⁵ Writing at a moment when statistics were closely associated with the physical sciences, these scholars revealed the field's surprising origins: in law, gambling, statecraft, and the human sciences, as well as astronomy and mathematics. Personal data, they explained, came in a rush, an “ava-

⁵ Lorraine Daston, *Classical Probability in the Enlightenment* (Princeton, N.J., 1988); Alain Desrosières, *The Politics of Large Numbers: A History of Statistical Reasoning* (Cambridge, Mass., 1998); Gerd Gigerenzer, Zeno Swijtink, Theodore Porter, Lorraine Daston, John Beatty, and Lorenz Krüger, *The Empire of Chance: How Probability Changed Science and Everyday Life* (Cambridge, 1989); Ian Hacking, *The Emergence of Probability: A Philosophical Study of Early Ideas about Probability, Induction, and Statistical Inference* (Cambridge, 1975); Hacking, *The Taming of Chance* (New York, 1990); Theodore M. Porter, *The Rise of Statistical Thinking 1820–1900* (Princeton, N.J., 1986); Stephen Stigler, *The History of Statistics: The Measurement of Uncertainty before 1900* (Cambridge, Mass., 1986). On independent statistics programs, see Alma Steingart, “Conditional Inequalities: American Pure and Applied Mathematics, 1940–1975” (PhD diss., MIT, 2013), 28–9.

lanche of printed numbers,” as Ian Hacking has memorably put it, and provided much grist for the early statistical mill starting in 1820–40 Europe. The avalanche marked the ascent of new bureaucracies given life by fears of social disorder and dreams of controlling subject or troubled populations, including Native Americans in the United States or potential revolutionaries in France and Germany.⁶

In the United States, the census of 1980, which provoked controversy over “under-counting” of minority groups, drew attention from mainstream historians to once rather technical topics, like the history of census taking or the evolution of vital registration systems, that had previously been most useful only as guides to the sources employed by early social historians.⁷ Leading works after 1980 turned their attention to questions of power: to the roots of censuses in subject colonies, to the inherently political nature of state-administered counts, and to the resistance to enumeration that plagued nineteenth-century counters.⁸ The influence of Foucauldian ideas of biopolitics and governmentality, turning attention to the capacity of quantification to make and manage new kinds of abstract populations or subjects, intersected with postcolonial perspectives to highlight the subtle forms of influence and coercion flowing through even such apparently trivial objects as a questionnaire or so unintimidating a person as a Civil War veteran-cum-1890s census enumerator.⁹

Having decided to look, historians realized that some sort of paper trail had once followed in one way or another a wide variety of Americans via the manuscript census (after 1850, when the census started recording individuals); church or government records of births, marriages, and deaths; mortgage records or cotton-picking quotas tied to chattel slaves; anthropological ledgers containing measures of dead Native Americans’ crania; asylum files for the mentally ill; school records for children; credit reporters’ catalogs of merchants’ characters; small town farmers’ account books tracking webs of debt; life insurers’ files on farmers’ family health histories; and federal pension files for Union soldiers.¹⁰ And then there were, of course, personal records like the family bible, which tracked in private homes the sort of genealogical data that Markus Friedrich discusses in this volume. Most such records would seem quite porous by

⁶ Ian Hacking, “Biopower and the Avalanche of Printed Numbers,” *Hum. Soc.* 5 (1982): 279–95; Oz Frankel, *States of Inquiry: Social Investigations and Print Culture in Nineteenth-Century Britain and the United States* (Baltimore, 2006).

⁷ On the census of 1980, see Margo J. Anderson, *The American Census: A Social History* (New Haven, Conn., 1988); William Alonso and Paul Starr, eds., *The Politics of Numbers* (New York, 1989). The most prominent earlier work is James Cassedy, *Demography in Early America: Beginnings of the Statistical Mind, 1600–1800* (Cambridge, Mass., 1969). On its use by social historians, see Michael Kammen, review of *Demography in Early America*, by James Cassedy, *J. Amer. Hist.* 57 (1970): 402–4.

⁸ Anderson, *American Census* (cit. n. 7); Patricia Cline Cohen, *A Calculating People: The Spread of Numeracy in Early America* (Chicago, 1983).

⁹ Geoffrey C. Bowker and Susan Leigh Star, *Sorting Things Out: Classification and Its Consequences* (Cambridge, Mass., 1999); Martha Hodes, “Fractions and Fictions in the United States Census of 1890,” in *Haunted by Empire*, ed. Ann Laura Stoler (Durham, N.C., 2006), 240–70.

¹⁰ Anderson, *American Census* (cit. n. 7), 36; Bonnie Martin, “Slavery’s Invisible Engine: Mortgaging Human Property,” *J. South. Hist.* 76 (2010): 817–66; Edward E. Baptist, *The Half Has Never Been Told: Slavery and the Making of American Capitalism* (New York, 2014), 131–43; Ann Fabian, *The Skull Collectors: Race, Science, and America’s Unburied Dead* (Chicago, 2010); Theodore M. Porter, “Funny Numbers,” *Cult. Unbound* 4 (2012): 585–98; Josh Lauer, “From Rumor to Written Record: Credit Reporting and the Invention of Financial Identity in Nineteenth-Century America,” *Tech. & Cult.* 49 (2008): 301–24; Winifred Barr Rothenberg, *From Market-Places to a Market Economy* (Chicago, 1992); Sharon Ann Murphy, *Investing in Life: Insurance in Antebellum America* (Baltimore, 2010); Drew Gilpin Faust, *This Republic of Suffering: Death and the American Civil War* (New York, 2008).

modern standards, and most have been lost to us today. Still, they demonstrate that the creation and gathering of personal data are anything but recent phenomena. Indeed, the evidence of long-established bureaucratic drives to track populations confronts us as we wander the streets, as Anton Tanter and Reuben Rose-Redwood have shown in their studies of house numbering in the United States and Europe: house numbers arose first to aid in governing households, both in the collection of taxes and in the maintenance of military rolls.¹¹

In more recent years, we have looked to the past with eyes made curious about older data-based technologies—like charts, maps, and even paper forms—by our encounters with computers, the Internet, data analysis software, and smartphones. Some scholars have looked to the past for inspiration, like the data visualization researcher Michael Friendly who has called attention to the era after the avalanche of printed numbers, from 1850 to 1910 (or so) as the “golden age of statistical graphics.”¹² Others have excavated deep histories for our current fascinations with geographic information system (GIS) mapping, the Internet “form,” or even the PDF.¹³ Slowly, then, we have been sketching personal data’s first act.

Let me trace its outlines. In the nineteenth century, states, corporations, and voluntary organizations (the NGOs of the past) began producing data doubles to meet their governing ambitions. Those doubles tended to be short lived, to travel only short distances, and to reflect or record the exercise of power, rather than serve as a source of power. Data aggregates derived from all those doubles mattered much more during the “golden age of statistical graphics.”¹⁴ Aggregations helped justify abstractions like the “nation,” and they created new categories to describe and class people.¹⁵

An example will add some texture to my sketch. Examine figure 1, a map illustrating the prevalence of hernia in the American north in the mid-nineteenth century at the time of the Civil War. The map’s story began in January 1864. That month the Provost Marshal General’s Bureau, which had been charged by Congress during the Civil War with administering a nationwide draft and overseeing military volunteers, created a Medical Branch under the supervision of Jedediah Hyde Baxter. That branch in turn built a network of surgeons who administered medical exams for draftees and volunteers. In the process of each exam, the surgeon, if he followed Baxter’s instructions, asked the by-then-naked draftee to “stand erect, place his heels together, and raise his hands vertically above his head, the backs together, in which position he was required to cough and make other expulsive movements, while the abdomen, the inguinal rings, and the scrotum were examined for hernia.”¹⁶ Examining surgeons in turn submitted detailed

¹¹ Anton Tanter, “Addressing the Houses: The Introduction of House Numbering in Europe,” *Histoire & Mesure* 24 (2009): 7–30, on 17–8; Reuben Rose-Redwood and Anton Tanter, “Introduction: Governmentality, House Numbering, and the Spatial History of the Modern City,” *Urban Hist.* 39 (2012): 607–13.

¹² Michael Friendly, “The Golden Age of Statistical Graphics,” *Statist. Sci.* 23 (2008): 502–35.

¹³ Susan Schulten, *Mapping the Nation: History and Cartography in Nineteenth-Century America* (Chicago, 2012); Ben Kafka, “Paperwork: The State of the Discipline,” *Book Hist.* 12 (2009): 340–53; Lisa Gitelman, *Paper Knowledge: Toward a Media History of Documents* (Durham, N.C., 2014).

¹⁴ Friendly, “Golden Age” (cit. n. 12).

¹⁵ Hacking, “Biopower” (cit. n. 6).

¹⁶ J. H. Baxter, *Statistics, Medical and Anthropological, of the Provost-Marshall-General’s Bureau, Derived from Records of the Examination for Military Service in the Armies of the United States during the Late War of the Rebellion, of over a Million Recruits, Drafted Men, Substitutes, and Enrolled Men* (Washington, D.C., 1875), 1:iv.

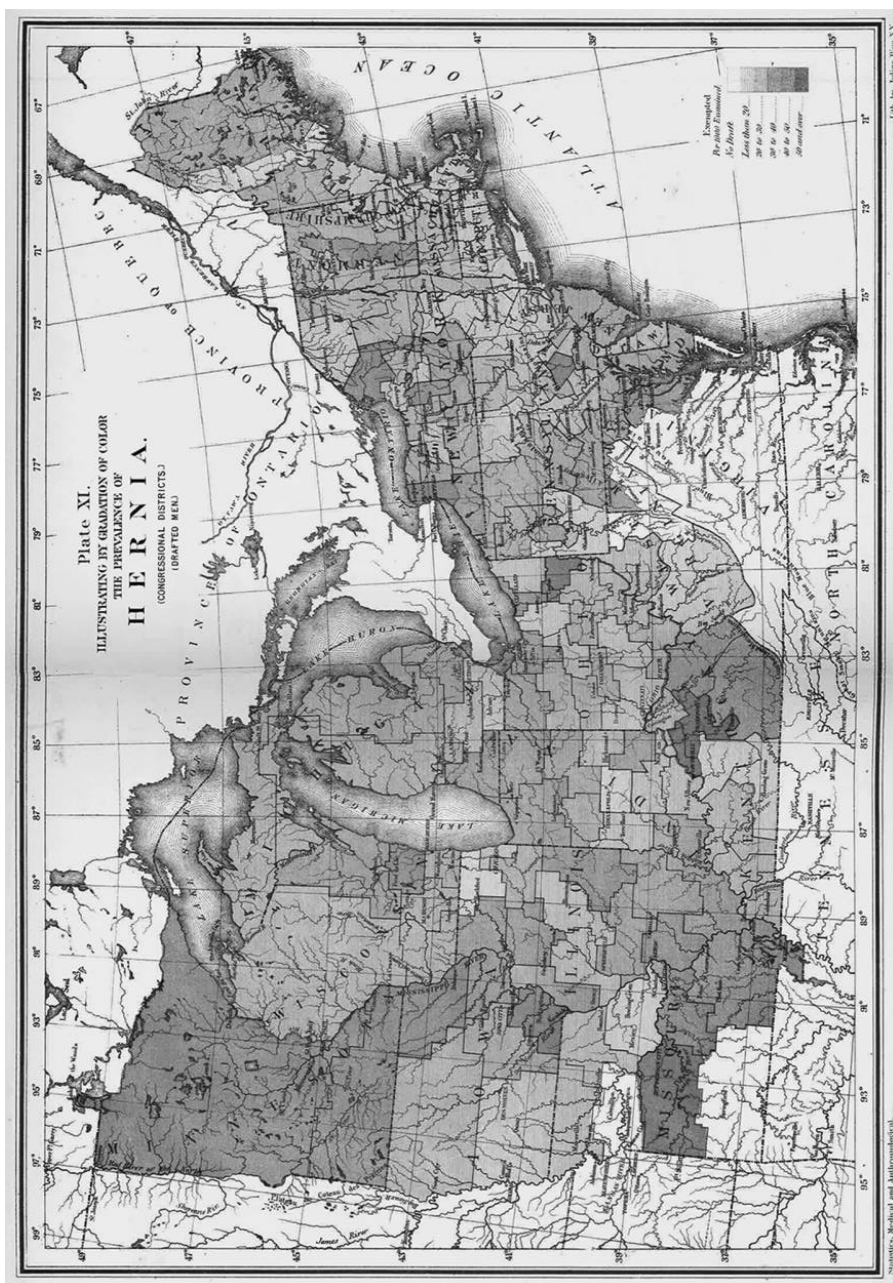


Figure 1. Baxter's map of hernia prevalence, derived from data collected during examinations of volunteers and draftees during the American Civil War. Baxter, *Statistics* (cit. n. 16), pl. xi.

records to Baxter after they accepted or rejected each candidate for military service. Baxter compiled these examination records and proceeded to make a series of statistical analyses.

Baxter's hernia map depended on blanks of paper filled out by surgeons for over half a million men. That meant half a million data doubles. Only one question on each paper really mattered to the individuals being examined and to the United States government's military apparatus: question 17 asked whether the examined individual was accepted or rejected and then, if the individual was rejected, queried the cause. The rest of the blanks were, in a sense, ancillary to the examination for war. They asked about age, nativity, occupation, "complexion," chest girth, "white or colored," and quite a few more factors, including eye color.¹⁷ The blanks reflect medical ideas of the time and evidence Baxter's keen interest in using the war examinations to conduct statistical investigations about the American population. This appears to have been the surgeons' understanding as well, since in the letters they wrote Baxter to accompany their doubles, the few who commented about the blank forms either requested a better system of disease classification "for statistical purposes," or complained about the time lost to filling out the blanks.¹⁸

Baxter's data doubles themselves had little power or value. They might have had some disciplining effect in keeping examining surgeons honest in their judgments. Yet the fact that Baxter corresponded extensively with those surgeons, asking their advice about procedural matters, suggests he trusted his examiners.¹⁹ The blanks mainly recorded a decision (reject or accept) that had already been made. They then went on to live secluded lives. Many blanks persist today in the National Archives, but they have been relegated to the Provost Marshal General's files, while many other individual service records were long ago integrated into individual files used to administer pensions or track Civil War service.²⁰ The short active life span and limited range of these data doubles appear typical of personal data's first act, just as it is typical that the data aggregates produced from those doubles proved powerful and traveled widely.

Baxter's blanks became much more significant when he brought them together in the 1870s for his statistical inquiries. Together, they enabled the creation of new ways of conceiving of the nation and of those within it. Baxter argued that his "statistical matter" related to "the people; the men engaged in every occupation; the professional man and the man of letters, the trader, the merchant, the clerk, the artisan, and the unskilled laborer; the rich man and poor man; the robust and the crippled; in short to the citizens of the United States, both native and foreign-born, and does, it is believed, illustrate the physical aptitude of the nation for military service."²¹ Baxter crafted an image of a nation, one that conceived of the United States in terms of a statistical population, in the manner championed by European thinkers like Adolphe Quetelet,

¹⁷ *Ibid.*, v.

¹⁸ *Ibid.*, 167, 345.

¹⁹ On quantifying as a "technology of distance" and as a means by which a bureaucracy disciplines its field agents, see Theodore M. Porter, *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life* (Princeton, N.J., 1995), ix.

²⁰ For discussion of the compiled military service records and of the Provost Marshal General File, see <http://www.archives.gov/research/military/civil-war/resources.html#discuss> (accessed 8 February 2017); <http://www.archives.gov/research/guide-fed-records/groups/110.html> (accessed 8 February 2017). On Civil War personal data, efforts to recover information about veterans, and the compiled military service records generally, see Faust, *Republic of Suffering* (cit. n. 10), 255.

²¹ Baxter, *Statistics* (cit. n. 16), vi.

presented as a map, echoing earlier Prussian examples, and set to depict a distinct set of social fears: here of a nation beset by disability and disease.²² His lithographer, Julius Bien, printed a series of similar and if anything even more elaborate maps and charts around the same time for the census director Francis Amasa Walker, whose maps “aided state governance by transforming people into an abstract ‘population’ with particular attributes that could be managed and administered,” according to historian Susan Schulten.²³

Baxter’s map structured social imaginings, blending medical knowledge with political boundaries to make the nation visible in a new way. In his reading of statistical work through a Foucauldian lens, Hacking explained statisticians’ creative potential to define social entities, a potential made real in every new category generated by enumerators numbering a population.²⁴ In Baxter’s map, this power manifested itself as a transformation of a political designation (the congressional district) into a medical object, with the possible transformation, in turn, of citizens’ perceptions of the places in which they lived.

Data aggregates could also exert power through the governing powers of states and localities, by making populations “legible,” to use James Scott’s terminology.²⁵ Placed in the hands of the burgeoning public health apparatus, such perceptions could become a reason for state intervention in daily affairs.²⁶ Local elites apparently understood the potential power of data aggregates even at the time of the surgeons’ examinations, and that understanding may have perverted Baxter’s collected hernia data by making some districts appear more herniated than they were. During the draft, Baxter explained, local quotas for recruits were set by comparing ratios of men serving to eligible men in each district. Hernia, he wrote, was an easily identified disqualifier, and so encouraging extra scrutiny for hernias became a likely means by which “local authorities, or even interested private individuals” could lower a locality’s quota.²⁷ So, as an aggregated mass, data doubles could exert power over the individuals they described, but that power came through group classifications and governing regulations.

Baxter’s maps and the rest of his data aggregates proved much more durable and mobile than the Provost Marshal General’s data doubles. Medical journals praised Baxter’s work and particularly its charts for the vision of America and its citizens they laid bare.²⁸ The book and maps were distributed widely, as the *New York Medical Journal* explained, by senators and congressmen to interested constituents.²⁹ Such mobility was typical.

Friendly’s graphical “golden age” qualified as such because bureaucracies with money had the will and resources to print nice lithographs, whether maps, charts, or diagrams, to accompany a new genre of social reform literature and statistical re-

²² See *ibid.*, lxxvii–lxxix; Schulten, *Mapping the Nation* (cit. n. 13), 164.

²³ Schulten, *Mapping the Nation* (cit. n. 13), 184.

²⁴ Hacking, “Biopower” (cit. n. 6).

²⁵ James C. Scott, *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed* (New Haven, Conn., 1999), 2.

²⁶ Charles E. Rosenberg, *The Cholera Years: The United States in 1832, 1849, and 1866* (Chicago, 1987).

²⁷ Baxter, *Statistics* (cit. n. 16), 88–9. This is a case of funny numbers. See Porter, “Funny Numbers” (cit. n. 10).

²⁸ See, e.g., *New York Med. J.* 24 (1876): 70–3, on 72. Schulten cites other reviews in *Mapping the Nation* (cit. n. 13), 182n54.

²⁹ *New York Med. J.* (cit. n. 28), 73.

porting.³⁰ The will to publish came out of a will to govern and control, to be sure. Hacking has pointed to the widespread belief that “one can improve—control—a deviant subpopulation by enumeration and classification.”³¹

But it also came out of a desire to display. It is more than a coincidence that the golden age bears the same dates as the craze for Victorian exhibitions and world fairs set off by the so-called Crystal Palace exhibition in London in 1851.³² When states, corporations, or reform organizations printed big, beautiful books or artful statistical graphics, they knew they could display their wares at the next exhibition, which was never very long in coming. The history of statistical display in the exhibition era remains to be written, and yet evidence from guidebooks, official descriptions, and printed catalogs suggests statistical artistry regularly on display across Europe and the Americas.³³ Large statistical presentations even showed up in parades.³⁴ States used statistical charts and books to showcase their modernity and their capacity to know their subjects, while corporations made similar moves to enhance their prestige. One prime case is the publication, paid for by seventeen American life insurers, of a volume on American vital statistics. The insurers printed it not only in New York, but also in London, Paris, and Madrid, apparently so they could brag at home of the reach of their influence, since they sold little insurance in Europe.³⁵

Having been fashioned either to tend to populations or to enhance prestige, data aggregates usually traveled within exchange economies driven by a spirit of collective scientific inquiry. When, for instance, the American Geographical and Statistical Society in New York tried to acquire statistical materials (still a rare commodity in the mid-nineteenth-century United States), it could buy some volumes, but it relied even more on donations and exchanges.³⁶ The group’s librarian reported accessions for a single month in 1859: “By donation, 200 volumes, 1,050 pamphlets, and 34 maps and prints; and by purchase 201 volumes and one globe.”³⁷ Many donated materials came from governments (some foreign, some national, some more local, as in the case of materials submitted by the mayor of Lawrence, Mass., about city schools and finances),³⁸ others from regional statistical or scientific societies like the Academy of Sciences in Stockholm or “a German natural history society,”³⁹ and others from private corporations and private authors.⁴⁰ Data doubles did not usually move in these networks.

³⁰ Frankel calls this “print statism” in *States of Inquiry* (cit. n. 6), 8–13.

³¹ Hacking, *Taming of Chance* (cit. n. 5), 3.

³² See Robert W. Rydell, *All the World’s a Fair: Visions of Empire at American International Expositions* (Chicago, 1984).

³³ For an example of a printed catalog, see Mutual Life Insurance Company of New York, *Introduction to the Exhibit of the Mutual Life Insurance Company of New York* (Paris, 1900); for references or images of statistical charts in displays, see, e.g., John D. Philbrick, *The Catalogue of the United States Collective Exhibition of Education* (London, 1878), 48; E. Monod, *L’Exposition Universelle de 1889* (Paris, 1890), 545.

³⁴ Willard C. Brinton, *Graphic Methods for Presenting Facts* (New York, 1914), 343.

³⁵ James Wynne, *Report on the Vital Statistics of the United States, Made to the Mutual Life Insurance Company of New York* (New York, 1857).

³⁶ Paul J. FitzPatrick, “Statistical Societies in the United States in the Nineteenth Century,” *Amer. Stat.* 11 (1957): 13–21.

³⁷ “Proceedings,” “Library Department,” *J. Amer. Geogr. Stat. Soc.* 1, no. 10 (1859): 287.

³⁸ “Library Department,” *J. Amer. Geogr. Stat. Soc.* 1, no. 7 (1859): 223.

³⁹ “Library Department,” *J. Amer. Geogr. Stat. Soc.* 1, no. 9 (1859): 285.

⁴⁰ “Library Department” (cit. n. 38), 223.

They normally remained local in scope, seldom traveled in exchanges, and were, except in the case of credit reporting, not made into commodities themselves. We might attribute data doubles' quiet lives to the same sort of fear of centralized power that stimulated the American Revolution, to nineteenth-century individuals who refused to be recorded, numbered, or tracked.⁴¹ Yet that cannot be the whole story, since those same revolutionaries were willing to allow themselves to be counted by the census in exchange for representation, and their near descendants proved open to bargaining away their personal data for valuable commodities like life insurance.⁴² More likely, data doubles may not have moved much because of the expense of moving them: as the economists might say, information, storage, and transport costs remained high.⁴³ To support this conclusion, we need only look to Prussia, which regularly sold its counting cards "as spoilage" once aggregations were completed, as Christine von Oertzen notes.⁴⁴

Two processes controlled the political economy of personal data in its first act. First, personal data provided the fodder necessary to structure emerging ideas of the "social" within and between nations. Maps and charts circulating in expositions and among bureaucrats provided tools for making sense of a nation's inhabitants in broad terms, using big categories like race, class, sex, state, and region and smaller categories related to health or occupation. They disseminated biopolitical tools and taught the means for making people into "populations," whether for the purpose of allocating them political power, providing them relief, or seeking reform. They also advertised the power of states, corporations, and reformers to intervene in the lives of the masses (or at least dream that they could), winning for their creators praise and prestige in an economy where both served as currency.

Second, personal data made possible new kinds of subjectivity. Cultural power flowed through data doubles, even if they were fleeting, local, and individually inconsequential. As bureaucracies came to fill their files with individual records, for instance, they participated in making the nineteenth century the "age of the first person singular," as Ralph Waldo Emerson dubbed it.⁴⁵ Data doubles in this period came to take individuals as their primary targets, rather than older units such as the household. The act of enumeration can, in and of itself, teach people to see themselves through the lens of new categories, as Martha Hodes has argued.⁴⁶ Individuals might therefore have come to see themselves, as individuals, through the options allowed on a mass-produced blank form. With the widespread publication of inscriptions based on aggregates, even more people—like the many who attended international expositions—

⁴¹ On fears of tyranny and corruption, see Bernard Bailyn, *The Ideological Origins of the American Revolution* (Cambridge, Mass., 1967), 48–54; on earlier concerns about censuses, see Cohen, *Calculating People* (cit. n. 8), 34–40, 47–80, 231n50.

⁴² "The Apportionment of Members among the States," *Federalist Papers*, no. 54 (1788); Dan Bouk, *How Our Days Became Numbered: Risk and the Rise of the Statistical Individual* (Chicago, 2015), xxiv.

⁴³ Naomi R. Lamoreaux, Daniel M. B. Raff, and Peter Temin, "Beyond Markets and Hierarchies: Toward a New Synthesis of American Business History," *Amer. Hist. Rev.* 108 (2003): 404–33, on 414.

⁴⁴ Christine von Oertzen, "Machineries of Data Power: Manual versus Mechanical Census Compilation in Nineteenth-Century Europe," in this volume.

⁴⁵ Louis P. Masur, "'Age of the First Person Singular': The Vocabulary of the Self in New England, 1780–1850," *J. Amer. Stud.* 25 (1991): 189–211, on 205–6.

⁴⁶ Hodes, "Fractions" (cit. n. 9), 256–8.

encountered social structures and categories that they could use to define themselves, as individuals.

ACT II: FITTING INTO THE MASS

In the first half of the twentieth century, a new adjective joined the abstraction “society,” the term that was a focus in personal data’s first act. Society became “mass,” and mass pops up throughout our twentieth-century histories explaining a series of fundamental transitions in ordinary life. In the United States, investment capital flowed into gigantic trusts around the turn of the century, facilitating a “great merger movement” that integrated industries horizontally and created monopolies or oligopolies. Those companies in turn integrated vertically, controlling within a single firm the entire supply chain and distribution system.⁴⁷ They brought to maturity mass production, mass distribution, and mass marketing.⁴⁸ With radio, nickelodeons, and new techniques in advertising, mass society came into being. Which is not to say that mass conformity took hold, or that all people truly had the same experiences.⁴⁹ But for those selling the same product to the large mass, it became convenient to identify or even construct uniformities. And so our dominant stories of personal data in this era highlight the collection of data—by social scientists, pollsters, and advertisers especially—meant to constitute and describe the “mass.”⁵⁰

The urge to understand the mass went hand in hand with a belief that universals trumped particularities in social and human science research, and that archives might in fact be filled to a point approaching total coverage.⁵¹ Existing historical accounts point to the early years of the Cold War as the apotheosis of the drive to know and measure everything, from natural phenomena like the aurora borealis, as Elena Aronova points out in this volume, to the most personal of data, like the dreams about which Lemov writes.⁵² A fantasy of preservation flourished at midcentury, whether in popular literature, such as Isaac Asimov’s *Foundation*, or in enthusiasms for time capsules and seed banks.⁵³

The principal urge of personal data’s second act was to fit individuals into the mass. Fitting could involve sorting and prediction—it could mean determining how a particular person related to and would in the future relate to the sort of social categories

⁴⁷ Naomi R. Lamoreaux, *The Great Merger Movement in American Business, 1895–1904* (Cambridge, 1988).

⁴⁸ Alfred D. Chandler Jr., *The Visible Hand: The Managerial Revolution in American Business* (Cambridge, Mass., 1977), pt. 3.

⁴⁹ Lizabeth Cohen, *Making a New Deal: Industrial Workers in Chicago, 1919–1939* (Cambridge, 1990), 120–47.

⁵⁰ Sarah E. Igo, *The Averaged American: Surveys, Citizens, and the Making of a Mass Public* (Cambridge, Mass., 2007); Olivier Zunz, *Why the American Century?* (Chicago, 1998).

⁵¹ This was a good moment, in other words, for “sciences of the archive.” See Lorraine Daston, “The Sciences of the Archive,” *Osiris* 27 (2012): 156–87.

⁵² See Elena Aronova, “Geophysical Datascares of the Cold War: Politics and Practices of the World Data Centers in the 1950s and 1960s,” in this volume; Lemov, “Anthropology’s Most Documented Man” (cit. n. 1). On the widespread gathering of personal data in the early Cold War, see also Lemov, “Filing the Total Human: Anthropological Archives from 1928 to 1963,” in *Social Knowledge in the Making*, ed. Charles Camic, Neil Gross, and Michelle Lamont (Chicago, 2011), 119–50; Joel Isaac, “Epistemic Design: Theory and Data in Harvard’s Department of Social Relations,” in *Cold War Social Science*, ed. Mark Solovey and Hamilton Cravens (New York, 2012), 79–95.

⁵³ Isaac Asimov, *Foundation* (New York, 1951); Jack Ralph Kloppenburg Jr., *First the Seed: The Political Economy of Plant Biotechnology, 1492–2000* (Cambridge, 1988), chap. 7.

created in the prior century. Fitting could also mean something more procrustean—it could mean managing and controlling individuals such that they acted more like the mass.⁵⁴ In many cases, the makers of personal data sought to fit their contemporaries to the mass mold in both ways at once.

For those building a mass society, it made sense to view individuals through statistical lenses. By making people into “statistical individuals,” it became possible to sort them according to the futures the statistics predicted they would have.⁵⁵ World War I created fertile bureaucratic soils for advocates of sorting and prediction, most notably in the field of mental testing, which promised to produce representations of thinking selves to aid officers in selecting and placing soldiers.⁵⁶ After the war, mental testing inspired other sorts of statistical individuals. Educational Testing Service (ETS), for instance, succeeded in making the SAT examination, which scored test takers in normalized percentiles, into a necessity for getting into college. Buoyed by success, ETS’s first president, Henry Chauncey, imagined designing a “Census of Abilities” capable through extensive testing of a “great sorting,” as Nicholas Lemann has described it.⁵⁷ In the meantime, social scientists in colleges and universities used mental tests to predict future success in corporate life, which drew the financial support of men like W. T. Grant, who used his department store wealth to fund the most famous such study at Harvard beginning in the 1930s.⁵⁸

Of course, one could control or manage men without making statistical individuals, as becomes clear when we consider the production of representations of working selves in the early twentieth century. The spread of “scientific management” and “time” and “motion” study, as propounded by F. Winslow Taylor and Frank Gilbreth, meant that data doubles came to fill company files, with papers describing the motions of workers in the machine shop or surgeons in their operating chambers, while standardized evaluation forms defined some workers as exemplary and others as in need of improvement.⁵⁹ In Henry Ford’s extreme case, workers’ private and working lives blended together in the records of the Ford Sociological Department, which was charged with keeping workers on the straight and Americanized narrow.⁶⁰ With the coming of automation, some individuals’ motions even became the pattern for a later machine’s operation.⁶¹ The novelist Kurt Vonnegut captured concerns over the appropriation and reproduction of a data double—so as to put the model for that double

⁵⁴ Beniger casts the urges to control and manage, even in what this essay would call the third act of personal data, in terms of a long “control revolution” beginning with the Industrial Revolution. See James R. Beniger, *The Control Revolution: Technological and Economic Origins of the Information Society* (Cambridge, Mass., 1986).

⁵⁵ Bouk, *Our Days* (cit. n. 42).

⁵⁶ John Carson, *The Measure of Merit: Talents, Intelligence, and Inequality in the French and American Republics, 1750–1940* (Princeton, N.J., 2006), chap. 6.

⁵⁷ Nicholas Lemann, “The Great Sorting,” *Atlantic Monthly* (September 1995): 84–100, on 88–90.

⁵⁸ Joshua Wolf Shenk, “What Makes Us Happy?,” *Atlantic Monthly* (June 2009): 36–53; Heather Munro Prescott, “Using the Student Body: College and University Students as Research Subjects in the United States during the Twentieth Century,” *J. Hist. Med.* 57 (2002): 3–38, on 19–21.

⁵⁹ Frank B. Gilbreth, *Primer of Scientific Management* (New York, 1912); W. E. Bloom, *Lincoln Factory Executive Service: Time, Motion, and Methods Study* (Cleveland, 1960); Caitjan Gainty, “‘Going after the High-Brows’: Frank Gilbreth and the Surgical Subject, 1912–1917,” *Representations* 118 (2012): 1–27.

⁶⁰ Stephen Meyer III, *The Five Dollar Day: Labor Management and Social Control in the Ford Motor Company, 1908–1921* (Albany, N.Y., 1981), chap. 6.

⁶¹ David F. Noble, *Forces of Production: A Social History of Industrial Automation* (New York, 1984).

out of work—in *Player Piano*: “Makes you feel kind of creepy, don’t it, Doctor, watching them keys go up and down? You can almost see a ghost sitting there playing his heart out.”⁶² In such data doubles, prediction mattered less than control. Managers studied existing work methods to find the “one best way” and recorded efficient workers’ actions so as to spur other laborers, or eventually machines, to work to the same standard.⁶³

Overlapping efforts to predict, sort, and manage produced a profusion of data doubles in personal data’s second act, each tied to a particular kind of self in the making. The Eugenics Record Office, for instance, set out to trace Americans’ heredity, inscribing on paper a eugenic self for about 35,000 people.⁶⁴ Colleges measured new entrants as part of physical education curricula or took “posture photos” of students as a step toward literally straightening America’s backbone, leaving behind records of embodied selves.⁶⁵ Consuming selves emerged in mail order companies’ files, in the 7.5 million customer records created by Sears, Roebuck & Company around 1920.⁶⁶ By 1938, the tabulating cards kept by the Social Security Board recorded the pay records of 40 million workers, even as the Internal Revenue Service described the taxable selves of a growing proportion of the population, over one-third of Americans in 1945.⁶⁷ In some cases, people crafted identities from these bureaucratic selves, like those who displayed their social security numbers as tattoos on their bodies.⁶⁸

Industrial consolidation beginning at the turn of the century followed by two world wars encouraged the growth of managerial apparatuses dependent on the constant flow of paper people. Those factors, as part of the larger process of making a mass society, made data doubles more valuable, while new technologies—first and foremost, the paper card, in conjunction with the card filing system, the tabulator, and the microcard—made data doubles more practical. If the nineteenth century was the era of the ledger, and our own times are dominated by computers and “smart” devices, the vast majority of the twentieth century belonged to the card file, to systems born of librarians’ organizing ambitions coincident with the search for new efficiencies by banks and insurers.⁶⁹

Life insurers were early adopters of card-filing technologies, having realized the capacity of cards to allow for wider sharing of data. Medical directors from the nation’s largest insurers banded together in 1902 to form the Medical Information Bureau (M.I.B.), which came into being as a series of paper cards (a complete set for each company, supplemented and updated through Melvil Dewey’s Library Bureau) con-

⁶² Kurt Vonnegut, *Player Piano* (1952; repr., New York, 2006), 32.

⁶³ Robert Kanigel, *The One Best Way: Frederick Winslow Taylor and the Enigma of Efficiency* (New York, 1997).

⁶⁴ Garland E. Allen, “The Eugenics Record Office at Cold Spring Harbor, 1910–1940: An Essay in Institutional History,” *Osiris* 2 (1986): 225–64, on 251.

⁶⁵ Ron Rosenbaum, “The Great Ivy League Nude Posture Photo Scandal,” *New York Times Magazine*, 15 January 1995.

⁶⁶ On mail order as a model site, second only to life insurance, for the use of statistics and personal data in business planning, see J. George Frederick, *Business Research and Statistics* (New York, 1920), 70–3.

⁶⁷ Arthur J. Altmeyer, “Three Years’ Progress toward Social Security,” *Soc. Security Bull.* 1 (1938): 1–7, on 1; Daniel J. Boorstin, *The Americans: The Democratic Experience* (New York, 1974), 209.

⁶⁸ Bouk, *Our Days* (cit. n. 42), 230–6.

⁶⁹ Gerri Flanzraich, “The Library Bureau and Office Technology,” *Libraries & Cult.* 28 (1993): 403–29; Markusz Krajewski, *Paper Machines: About Cards and Catalogs, 1548–1929* (Cambridge, Mass., 2011).

taining data on any known impairments discovered during participating companies' routine medical examinations of applicants. The card exchange helped life insurers ferret out potential fraud or medical examiner error and served as a basis for decisions to deny applicants insurance or to issue a substandard policy charging higher premiums. The companies, recognizing that sharing sensitive medical data was not likely to be popular, forbade medical directors from talking about M.I.B. even with their own medical examiners, agents, or managers.⁷⁰ Because of the sensitivity of these data doubles, combined with their value to medical directors hoping to better select long-lived applicants, they were well guarded. One medical director aptly explained that his company kept its cards "behind iron doors" to restrict access to a few competent "medical men." Another described a veritable gauntlet of gates and watchful clerks guarding access to the M.I.B. card files.⁷¹ Companies valued these data doubles highly enough to grant them their own safes.⁷² Yet even as they shared cards within their private club, M.I.B. members resisted the commodification of their data doubles. When the American Life Convention, an organization composed mostly of smaller, more recently established firms, tried to create a similar exchange and asked to purchase access to some M.I.B. files, the M.I.B. refused the request and used its data power to lure the larger American Life Convention companies into the M.I.B. stables.⁷³

Sharing data on cards also facilitated the mass production of statistical individuals. Consider the tabulating card pictured in figure 2. Life insurers used these cards in a 1912 study that collected data on individuals (identified by policy numbers) in order to determine which factors could best be used to predict longevity. The card's categories suggest all the possible factors, from occupation to the term of the life insurance policy purchased to the kind of "impairments" noted by examining doctors at the time of application. A team of life insurance medical directors and actuaries used the study's findings to improve a numerical rating system that could partially automate decisions about whom to grant a policy and that policy's terms. The tiny column "BLD," which stood for "build" and derived from a ratio of weight to height, held the data that served with age as the baseline for the rating systems. These cards, these data doubles, helped create a new data aggregate, as with Baxter's maps. But in this case, a bureaucracy already existed that could immediately take advantage of that new statistical finding. Many life insurance companies' applicants' data doubles would soon be interpreted using the results of the 1912 study. In viewing millions of doubles through the aggregate, insurers made statistical individuals from a significant proportion of the American population.

The second act's characteristic mobilization of data doubles to organize and control became evident in its changing modes of statistical visualization. An AT&T statistician named M. C. Rorty reflected on early efforts to reform statistical graphics and codified

⁷⁰ "Record of M. I. B. In the Proceedings of the Association of Life Insurance Medical Directors from May 1892- to December 1916," 7-8, in Folder "Medical Information Bureau 1916," Box RG/13-Subject Files M11, MetLife Archives. On the M.I.B., see Bouk, *Our Days* (cit. n. 42), 78-86.

⁷¹ Dr. Dwight, speaking at 7 October 1914 executive meeting, in "Record of M. I. B. in the Proceedings of the Association of Life Insurance Medical Directors from May 1892- to December 1916" (cit. n. 70), 38.

⁷² Frank Wells, speaking at 7 October 1914 executive meeting, in "Record of M. I. B. in the Proceedings of the Association of Life Insurance Medical Directors from May 1892- to December 1916" (cit. n. 70), 38.

⁷³ R. Carlyle Buley, *The American Life Convention 1906-1952: A Study in the History of Life Insurance* (New York, 1953), 1:440, 451-2, 466-7.

MEDICO ACTUARIAL MORTALITY INVESTIGATION																						
10 30		Year Issue		Duration		Cause of Death		Occupation		12		IMPAIRMENTS			Y		Kind of Policy		Amount of Insurance		Habitat	
X 2040				X		11		X		11		1st 2d 3d			K A B C D E F G H J		X X		X		X X	
						</																

a new set of graphical norms in a set of articles in the journal *Industrial Management* in 1920. Energized by wartime experience in logistics, Rorty noted the new importance for the industrial executive of readily understandable data. He explained, “even the most thoroughly trained executives, backed up by the best systems of records and accounts, will be aided very greatly in carrying the load of administration if masses of figures are made easy of comprehension by reduction to graphical form.”⁷⁴ Popular audiences still mattered, even if the world’s fairs took a break after World War I, but graphics intended to reach wider audiences tended to be less baroque with fewer ornate maps and many more line graphs and tables (and the occasional cartoon), at least in part because rather than being printed on fine paper in leather bindings, many of the new style graphics were slated to be mass printed and distributed widely, as was the case with Metropolitan Life Insurance Company’s *Statistical Bulletin*.⁷⁵

Rorty was one of a new generation of administrative statisticians, about whom we still know too little, tasked with making possible “statistical control” of products and people. Textbooks written to teach this new breed of professional underline the value ascribed to data collection as an end in itself. “Millions of dollars a year are spent in the collection of data,” opened one such text.⁷⁶ Another, in its distribution of topics, made clear that the collection of data would often have to be contracted and paid for. While some enterprises created data doubles as a matter of course, many more needed to hire surveyors, market researchers, or in-house statisticians to actively gather personal data, sometimes with a hefty charge attached.⁷⁷ Frederick L. Hoffman, a life insurance statistician, was like many of his contemporaries in thinking that what we might now call the “data exhaust” of his industry (life insurance) was not enough for effective administration. He built a library with over a hundred thousand resources, while taking regular rambles through Southern cemeteries typing out personal data (birth and death dates) from each gravestone.⁷⁸ Storing and maintaining collected personal data added further expense. As Rorty warned his zealous statistical brethren, “it will not do to forget the man who said that his card catalogues were working out beautifully, but he was afraid he would have to give up his business in order to keep them up to date.”⁷⁹

While World War I stimulated interest in the production and analysis of statistical individuals and spread the gospel of statistical control widely, World War II provides dramatic examples of how much ideas about the value and power of data doubles changed in the twentieth century. It is now widely understood that census data and registration data, whether gathered by the Nazis or by occupied governments, facilitated the Holocaust.⁸⁰ In such cases, military officials gained access to data doubles and used them to find and capture the individuals those doubles represented. In the United States, it does not appear that those responsible for interning Americans of Japanese ancestry had access to data doubles directly. However, historians have discovered that census officials cooperated, avidly, with the internment process, using

⁷⁴ M. C. Rorty, “Making Statistics Talk,” *Indust. Manage.* 60 (1920): 394–8, on 394–5.

⁷⁵ To see both the older and newer, spare style, see images in Brinton, *Graphic Methods* (cit. n. 34).

⁷⁶ *Ibid.*, 1.

⁷⁷ Frederick, *Business Research* (cit. n. 66), 96–7, 101.

⁷⁸ Bouk, *Our Days* (cit. n. 42), 116–22, 145.

⁷⁹ Rorty, “Making Statistics Talk” (cit. n. 74), 396.

⁸⁰ Götz Aly and Karl Heinz Roth, *The Nazi Census: Identification and Control in the Third Reich* (Philadelphia, 2004).

existing census rolls (and the data doubles therein) to generate new inscriptions telling officials how many Japanese Americans lived in a given census block, for instance.⁸¹ Where the Prussian government had once used data doubles as waste paper after census activities, such bits of paper came in the mid-twentieth century to be seen as governing tools that could not only create social categories as before but could also play active roles in finding, disciplining, and destroying individuals. I argue that such a revaluation of the data double was, in fact, the norm in personal data's second act.

Changing modes of resistance further illumine the shifting political economy of personal data. Toward the end of the first act, to take one example, African Americans in the 1880s and 1890s successfully pushed state governments to outlaw the use of race in predicting life spans and pricing insurance.⁸² Resistance came as a rebellion against a social category given power over a population's economic opportunities. In contrast, personal data's second act closed only a little after the so-called Free Speech Movement, led by students at the University of California, Berkeley, took as one of its targets the IBM punch card and its warning not to "fold, spindle, or mutilate."⁸³ The data double, rather than the social structure, became the object of protest.

ACT III: EXALTING AND EXPLOITING THE INDIVIDUAL

Personal data's third act has yet to fully play out. Still, we can draw tentative conclusions about its origins, trace its basic characteristics, and mull over its potentially revolutionary implications. The third act began in the 1970s amid economic crisis, technological change, and liberatory politics. At its center, we find the individual represented by increasingly persistent, interconnected data doubles or sometimes by (seemingly oxymoronic) personalized aggregations. We find the exaltation of new possibilities for individuals alongside growing concerns about the extraction, commoditization, and capitalization of personal data in a manner that amplifies inequalities and could imperil the idea of the liberal subject that data's first act served to establish.

What world wars were to the second act, economic crises were to personal data's third act. The global crises of the 1970s began a series of structural changes in the American (and global) economy that at once shifted capital toward services and finance and away from manufacturing,⁸⁴ intensified efforts to cut advertising costs and increase sales by targeting customers more carefully,⁸⁵ cut loose a generation of physicists and mathematicians set to work on government projects or at universities and sent them to work in corporate data analysis instead,⁸⁶ and limited funding from private and public sources for the production of new data while creating incentives for corporate raiders and cost-conscious boards of directors to take data left as a by-product

⁸¹ William Seltzer and Margo Anderson, "The Dark Side of Numbers: The Role of Population Data Systems in Human Rights Abuses," *Soc. Res.* 68 (2001): 481–513, on 492, 498.

⁸² Bouk, *Our Days* (cit. n. 42), 41–8.

⁸³ Steven Lubar, "'Do Not Fold, Spindle or Mutilate': A Cultural History of the Punch Card," *J. Amer. Cult.* 15 (1992): 43–55, on 45–6.

⁸⁴ Greta R. Krippner, "The Financialization of the American Economy," *Socio-Econ. Rev.* 3 (2005): 173–208.

⁸⁵ Joseph Turow, *Breaking up America: Advertisers and the New Media World* (Chicago, 1997), chap. 6.

⁸⁶ David Kaiser, "Booms, Busts, and the World of Ideas: Enrollment Pressures and the Challenge of Specialization," *Osiris* 27 (2012): 276–302, on 285, 298; Emanuel Derman, *Models Behaving Badly: Why Confusing Illusion with Reality Can Lead to Disaster, on Wall Street and in Life* (New York, 2011).

of other activities out of company magnetic tapes and hard disks and market it directly.⁸⁷ What had been seen as boring business segments—from banking to media buying to data analysis—came to be seen as among the most creative.⁸⁸ Customer loyalty programs, which offer benefits such as discounted prices at the grocery store or rewards for frequent flyers, soon looked like so many golden geese because they also produced extensive data doubles that advertisers, market researchers, and political groups were willing to pay for.⁸⁹ Media analysis and credit reporting companies, like A. C. Nielsen and TRW, expanded their business, becoming brokers of data doubles, some gleaned from government sources (the remnants of personal data's first act) and others from new market sources.⁹⁰

Data brokers could reproduce, sell, and transfer data doubles more easily as digital, networked storage displaced card systems and ledgers. In the 1960s, IBM released the System/360 line of computers, which in turn became standard tools for large organizations.⁹¹ Around that same time, work with military projects, building on corporate efforts to more effectively keep track of their workers and materials, led to the creation of the “database” concept and software for manipulating large volumes of data.⁹² Networking new and more powerful computers soon allowed for real-time tracking of personal data, most notably with the creation of airline reservation systems in the mid-1960s.⁹³ Personal data might have found a centralized home in the federal government, but plans for a National Data Center drew controversy that killed the chance of such a project. Personal data would instead reside, in many pieces, in many different (public and private) databases.⁹⁴

At the same time, power and potential resided more in data doubles. Demand for commoditized data doubles flowed from institutions bent on turning doubles into predictions. Highly engineered “risk” systems built on developments of the “actuarial age” that historian Caley Horan has identified in postwar America.⁹⁵ Risk systems promised to extend lives, improve justice, and expand credit. They generated personalized risk scores that located in the individual such disparate possibilities as heart

⁸⁷ Wendy Nelson Espeland and Paul M. Hirsch, “Ownership Changes, Accounting Practice and the Redefinition of the Corporation,” *Accounting Org. Soc.* 15 (1990): 77–96; Philip Mirowski, *Science-Mart: Privatizing American Science* (Cambridge, Mass., 2011), 15–6; Doogab Yi, “Who Owns What? Private Ownership and Public Interest in Recombinant DNA Technology in the 1970s,” *Isis* 102 (2011): 446–74.

⁸⁸ This shift is illustrated for banking, media buying, and data analysis, respectively, by Scott Patterson, *The Quants: How a New Breed of Math Whizzes Conquered Wall Street and Nearly Destroyed It* (New York, 2010); Joseph Turow, *The Daily You: How the New Advertising Industry Is Defining Your Identity and Your Worth* (New Haven, Conn. 2011); Michael Lewis, *Moneyball: The Art of Winning an Unfair Game* (New York, 2003).

⁸⁹ Turow, *Breaking up America* (cit. n. 85), 138–44.

⁹⁰ Paul Starr and Ross Corson, “Who Will Have the Numbers? The Rise of the Statistical Services Industry and the Politics of Public Data,” in Alonso and Starr, *The Politics of Numbers* (cit. n. 7), 415–47, on 419.

⁹¹ Martin Campbell-Kelly and William Aspray, *Computer: A History of the Information Machine* (New York, 1996), 144–7.

⁹² Thomas Haigh, “How Data Got Its Base: Information Storage Software in the 1950s and 1960s,” *IEEE Ann. Hist. Comput.* 31 (2009): 6–25; Haigh, “Charles W. Bachman: Database Software Pioneer,” *IEEE Ann. Hist. Comput.* 33 (2011): 70–80.

⁹³ Campbell-Kelly and Aspray, *Computer* (cit. n. 91), 174–5.

⁹⁴ Simson Garfinkel, *Database Nation: The Death of Privacy in the 21st Century* (Cambridge, 2000), 16.

⁹⁵ Caley Horan, “Actuarial Age: Insurance and the Emergence of Neoliberalism in the Postwar United States” (PhD diss., Univ. of Minnesota, 2011).

disease (via the risk factor popularized in the 1960s and 1970s), criminal activity (with many states tying parole to actuarial scoring systems from the 1970s on), or credit default (with FICO scores developed in the late 1970s).⁹⁶ The resulting proliferation of statistical individuals turned attention away from systemic or social causes for disease, recidivism, or credit default. Individuals were supposed to have gained the power—and responsibility—to heal themselves. Many commentators invoke a political project of “neoliberalism” to explain this trend, while Daniel T. Rodgers has argued that the shift away from social explanations with an accompanying skepticism of larger social categories (like race or gender) in favor of a focus on individual choice-making agents with fluid identities and weak social ties characterized an “age of fracture” that transcended political movements.⁹⁷

What had been merely a demand for data doubles came to look more like a new logic of extraction and accumulation with the growth of the Internet and of the tools we now associate with Big Data. New firms found in the Internet the capacity to make money by simultaneously commoditizing personal data while using such data internally in the style of old-school statistical control. Facebook exemplified that blending. It built a trove of customer data, harvested at little cost from its legions of users. Whether they acknowledged the deal or not, Facebook users traded their data for the right to craft an online persona, make connections, and see pictures of their “friends.” Facebook may not have sold that data directly, but it did monetize that data. It sold, and sells, finely tuned profiles to advertisers so they can target ads according to the peculiarities of each data double (and presumably of the person that data double represents).

Shoshana Zuboff has argued that a new form of capitalism is in the making. Surveillance capitalism works by hoarding user data to produce “behavioral surplus” (here Facebook’s data doubles), a form of capital gleaned from users that the capitalist can then use to produce “prediction products” for sale to third parties, like advertisers or insurers.⁹⁸ Facebook’s annual reports do not list the value of its data doubles—its surveillance capital—directly, but we can infer that value by looking at the company’s revenue per user, which in 2014 averaged \$9.45 worldwide, and quite a bit more for just the United States and Canada.⁹⁹ When the Organisation for Economic Co-operation and Development (OECD) wanted to determine the dollar value of personal data in 2010, it presented a range of possible metrics, from the cost to companies of having personal data stolen (around \$1.70 per person) to the cost of purchasing data from brokers (an average of \$35 for a robust record) to the outlying value implied by the cost of identity theft insurance (\$155).¹⁰⁰ Many marvel at how apparently valueless “digital

⁹⁶ William G. Rothstein, *Public Health and the Risk Factor* (Rochester, N.Y., 2008), 359–67; Robert A. Aronowitz, *Making Sense of Illness: Science, Society, and Disease* (New York, 1998), chap. 5; Bernard E. Harcourt, *Against Prediction: Profiling, Policing, and Punishing in an Actuarial Age* (Chicago, 2006), 40–1; Martha Poon, “Scorecards as Devices for Consumer Credit: The Case of Fair, Isaac & Company Incorporated,” in “Market Devices,” suppl. 2, *Sociol. Rev.* 55 (2007): 284–306, on 294–5.

⁹⁷ Daniel T. Rodgers, *Age of Fracture* (Cambridge, Mass., 2012).

⁹⁸ Shoshana Zuboff, “Big Other: Surveillance Capitalism and the Prospects of an Information Civilization,” *J. Inform. Technol.* 30 (2015): 75–89; Zuboff, “The Secrets of Surveillance Capitalism,” *Frankfurter Allgemeine Zeitung*, 5 March 2016.

⁹⁹ Facebook, *2014 Annual Report*, 29 January 2015, 37–8, <http://investor.fb.com/secfiling.cfm?filingID=1326801-15-6&CIK=1326801> (accessed 9 February 2015).

¹⁰⁰ OECD, “Exploring the Economics of Personal Data: A Survey of Methodologies for Measuring Monetary Value,” *OECD Digital Economy Papers*, no. 220 (2013), 5, <http://dx.doi.org/10.1787/5k486qtxldmq-en> (accessed 8 February 2017).

exhaust” can lead to such riches, while Zuboff argues that the label of “exhaust” is itself a way of hiding the inherent value in data doubles and justifying its extraction in acts of “dispossession by surveillance.”¹⁰¹

The prediction products that make money for Facebook or Google, while answering our questions or filling our newsfeeds, can be understood as algorithmic aggregations that themselves sometimes resemble data doubles. The geographer Louise Amoore explains how state security apparatuses produce these aggregates, too: scanning millions of travelers for possible terrorists, they follow retail data miners in treating much personal data as essentially disposable, focusing instead on the search for niche correlations. Linking together a series of correlations, new-style security creates a data aggregation capable, by definition, of encompassing only a very small population, or even just a handful of individuals.¹⁰² The quest for individual improvement through self-quantification might be thought of as carrying this trend to its logical conclusion.¹⁰³ Bernard Harcourt identifies a “doppelgänger logic” at the heart of Big Data, as algorithms search for “our perfect match, our twin, our look-alike, in order to determine the next book we want to buy . . . the perfect answer to our question.”¹⁰⁴ By this logic, the best aggregation mirrors the data double so that the individual can be simultaneously exalted and exploited.

Enthusiasts in personal data’s third act see a world of new possibilities dawning. Big data, exclaimed one influential book, is “poised to shake up everything from businesses and the sciences to healthcare, government, education, economics, the humanities, and every other aspect of society.”¹⁰⁵ Big Data, some claimed, could empower individuals to find and understand love, uncover new self-knowledge, or unleash the democratic yearnings of oppressed masses.¹⁰⁶ The 2011 film *Moneyball* encapsulates a popular vision for that shake-up. Its lead character, the general manager of the Oakland Athletics baseball team, Billy Beane, sets out a problem of inequality early in the film: “There are rich teams and there are poor teams.” An ensuing montage poses the solution, as grainy, digitized images of players flash across a monitor, alongside tables of data describing every aspect of their on-field performance. “This is a code that I’ve written for our year-to-year projections. This is building in all the intelligence that we have to project players,” explains Beane’s data-loving assistant. “It’s about getting things down to one number, using the stats the way we read them, we’ll find value players that nobody else can see. People are overlooked for a variety of biased reasons and perceived flaws: age, appearance, personality. Bill James and mathematics

¹⁰¹ Zuboff, “Secrets” (cit. n. 98).

¹⁰² Louise Amoore, “On the Emergence of a Security Risk Calculus for Our Times,” *Theory Cult. Soc.* 25 (2011): 24–43, on 36. On concerns about similar methods in the retail sphere, see Ryan Calo, “Digital Market Manipulation,” *George Washington Law Rev.* 82 (2014): 995–1051.

¹⁰³ Gary Wolf, “The Data-Driven Life,” *New York Times Magazine*, 28 April 2010; Natasha Dow Schüll, “Data for Life: Wearable Technology and the Design of Self-Care,” *Biosocieties* 11 (2016): 317–33.

¹⁰⁴ Bernard Harcourt, *Exposed: Desire and Disobedience in the Digital Age* (Cambridge, Mass., 2015).

¹⁰⁵ Viktor Mayer-Schönberger and Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think* (New York, 2013), 11.

¹⁰⁶ Christian Rudder, *Dataclysm: Who We Are (When We Think No One’s Looking)* (New York, 2014); on excitement (and subsequent disillusion) about Big Data’s potential for supporting social movements, see Zeynep Tufekci, “Engineering the Public: Big Data, Surveillance and Computational Politics,” *First Monday* 19 (2014), <http://firstmonday.org/ojs/index.php/fm/article/view/4901/4097> (accessed 8 February 2017).

cuts straight through that.”¹⁰⁷ Armed with data and with the algorithms to make sense of such data, Beane humiliates the team’s manager, centralizes power in the executive, and gets offered a big payout in the end, while his lovable band of misfit players do get (lower-wage) opportunities that they deserve. It is a parable: personal data will save us. Inequality can be solved by data put in the hands of a technocratic, centralized elite who will look past older social structures and past old prejudices to bring new success, and who will be paid well for their efforts.

The third act of personal data has not in fact transcended old social biases, however, or solved all problems of inequality. The language and spirit of the civil rights movement did make data-intensive companies and their regulators skeptical of traditional categories and practices, while a movement to win equality for women set in motion new antidiscrimination laws, notably the 1974 Equal Credit Opportunity Act, which eventually extended the promise of nondiscrimination to a host of “protected categories,” including gender, race, religion, age, and national origin.¹⁰⁸ New algorithms promised a formal indifference to those categories that would still allow for profitable discrimination (and so an increase in some forms of inequality), which we must understand as the essence of many Big Data projects. Auto insurers, for instance, have enthused over the possibility of “telematics”: devices that precisely monitor individual driving behavior, to provide more perfectly individualized metrics (how long one drives, how fast, how often one stops short, etc.) to be used in setting premiums.¹⁰⁹ Critics maintain, however, that data mining may still hurt protected groups defined in personal data’s first act and thus run afoul of the antidiscrimination laws that data mining was supposed to avoid.¹¹⁰ Indeed, a recent investigation of an algorithm used in many states to inform decisions about bail or sentencing discovered racial bias.¹¹¹ Moreover, indifference to legacies of injustice could be a hindrance in and of itself to justice.¹¹²

Critics fear that Big Data could even displace the liberal self and social order that took its modern form in personal data’s first act. Some worry about a corrosive “computational politics,” that the drive toward personalizing media could destroy the public sphere or at least exaggerate political, social, and economic inequalities within it.¹¹³ They worry that the “liberal ideal . . . no longer has traction in a world in which commerce cannot be distinguished from governing or policing or surveilling or just simply living privately.”¹¹⁴ They worry that Big Data “threatens the existential and

¹⁰⁷ *Moneyball*, directed by Bennett Miller (Los Angeles, 2011), DVD.

¹⁰⁸ Mary L. Heen, “Ending Jim Crow Life Insurance Rates,” *Northwest. J. Law & Soc. Policy* 4 (2009): 360–99; Louis Hyman, *Debtor Nation: The History of America in Red Ink* (Princeton, N.J., 2011), chap. 6.

¹⁰⁹ Bill Franks, *Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics* (Hoboken, N.J., 2012), 54–5.

¹¹⁰ Solon Barocas and Andrew D. Selbst, “Big Data’s Disparate Impact,” *Calif. Law Rev.* 104 (2016): 671–732; Danielle Keats Citron and Frank Pasquale, “The Scored Society: Due Process for Automated Predictions,” *Washington Law Rev.* 89 (2014): 1–33.

¹¹¹ 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, <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> (accessed 8 February 2017).

¹¹² Jonathan Simon, “The Ideological Effects of Actuarial Practices,” *Law Soc. Rev.* 22 (1988): 771–800.

¹¹³ Tufekci, “Engineering the Public” (cit. n. 106); Turow, *The Daily You* (cit. n. 88).

¹¹⁴ Harcourt, *Exposed* (cit. n. 104), 26.

political canon of the modern liberal order defined by principles of self-determination.”¹¹⁵ They wonder whether we are all on the verge of being reduced to servants of an engineered system fueled by our own data.¹¹⁶ If personal data served in building the liberal self and our ideas of society, might it also help to dismantle them?

CONCLUSION

This essay has set out a tripartite scheme describing the changing political economy of personal data, a scheme that tracks in unsurprising ways with recent attempts to describe broad changes in America’s broader political economy.¹¹⁷ In broad strokes, it might well describe transitions in the political economy of personal data in many nations. Beginning in the early nineteenth century and lasting through the first decade of the twentieth, states and corporations bent on ordering the wide world they hoped to rule made aggregates and fashioned new social categories that characterized the first act of personal data. World war and unprecedented corporate expansion brought the second act of personal data into being by focusing more on the use of data doubles, often interpreted through the lens of older social aggregations, to manage and control individuals. In our own act—personal data’s third, which began in the 1970s—data doubles became more powerful and more valuable, while the exaltation of the individual obscured both the lingering influence of older social categories and the capitalist fantasies fulfilled by flitting, fleeting “Big Data” aggregates, which are no less imposing for their evanescence.

¹¹⁵ Zuboff, “Secrets” (cit. n. 98).

¹¹⁶ Martha Poon, “Corporate Capitalism and the Growing Power of Big Data,” *Sci. Technol. & Human Values* 41 (2016): 1088–1108.

¹¹⁷ Lamoreaux, Raff, and Temin, “Beyond Markets” (cit. n. 43); Jonathan Levy, “Accounting for Profit and the History of Capital,” *Crit. Hist. Stud.* 1 (2014): 171–214.

link to
idea of
resisting
risk power
from within
→ 12
Koopman