Galton's Pride & Du Bois's Pursuit:

The Resilience of Data-Driven Inequality

DRAFT – PLEASE DO NOT CIRCULATE

Note: if printing, please first be aware of images at end of file on last three pages.

Colin Koopman

University of Oregon, Dept. of Philosophy

Abstract (from WPSA): Francis Galton is today remembered as an eminent prince of science if not also as a wild crank who, at the end of the nineteenth century, helped spawn the grandiose political disaster of eugenics. But before he was anything else, and beneath all he would do, Galton was first and foremost a tinkering technician of measure. There are numerous domains of science over which Galtonian conceptions of measure retain considerable influence. Among these are efforts in contemporary data science, which as critical data studies scholars Ruha Benjamin and Virginia Eubanks have recently shown, risk a bevy of racial injustices. Even where contemporary sciences officially disclaim the ideologies of racism that funded Galton's eugenics, they maintain such of his technologies as an informatics of race and a focus on the family. A viable alternative for the present can be found in the data portraits developed by Galton's contemporary, W.E.B. Du Bois.

§

Francis Galton is today remembered as an eminent prince of science if not also as a wild crank who, at the end of the nineteenth century, helped spawn the grandiose political disaster of eugenics, the ripples of which would continue to be felt far into the twentieth century. Moreso than prince or crank, Galton was first and foremost a tinkering technician of measure. He was engineer of a variety of means of tallying and quantifying, if not also at times a major theorist of all that measure adds up to. All of Galton's projects, from his scientific contributions to statistics to his politics of eugenics, depended on technologies of measure. Stephen Jay Gould observes that, "Quantification was Galton's god, and a strong belief in the inheritance of nearly everything he could measure stood at the right hand."[1] Galton himself was equally explicit: "until the phenomena of any branch of knowledge have been subjected to measurement and number, it cannot assume the status and dignity of a science."[2]

In his 1992 essay "Galton's Regret" the anthropologist of science Paul Rabinow considers Galton's own dissatisfaction at his attempts to leverage his beloved measures toward the scientific implementation of racism. His project in eugenics, essentially an attempt to give racism the inflection of biological and statistical science, certainly persisted even beyond his death. Yet it never really gained the scientific stature which Galton had hoped for it. Rabinow observes that this failure "constituted for the Victorian founder of eugenics a major disappointment."[3] Galton's regret.

Rabinow also observes how Galton's unrealized dream persists in ongoing efforts to articulate the high-tech bio-sciences of the twentieth century, namely genetics and genomics, to "older cultural understandings of race, gender, and age."[4] Rabinow's concern was that when we

---

[1] Gould 1981, 108
xxxGould, Mismeasure book.
[2] Galton 1879, 149
xxxGalton, Francis. 1879. "Psychometric Experiments" in *Brain* 2, 1879: 149-162.
[3] Rabinow 1992, 114
xxxRabinow, Paul. 1992. "Galton's Regret" in *Essays on the Anthropology of Reason*. Princeton: Princeton UP, xxx.
[4] Rabinow 1992, 127

translate the bio-sciences onto cultural configurations steeped in long histories of inequality, we risk producing new iterations of those inequalities even where we explicitly want to not be doing that. I share Rabinow's concern. In returning to Galton once again today we can see how this remains salient across an expanding field of scientific and cultural practices. I here seek to animate Rabinow's concern for other domains of data-driven science extending beyond the pre-eminent bio-sciences of genetics and genomics.[5]

Galton is eminently approachable as a figure in the history of our present, to draw on a methodological term of art of Michel Foucault's, whose work has been a serious influence for Rabinow and also for myself.[6] There are numerous domains of science over which Galtonian conceptions of measure retain considerable influence. Not least among these are a host of efforts in contemporary data science deeply reliant upon the kind of informational infrastructure Galton helped develop. Projects in human measure have recently reassumed pride of place in proving themselves integral to twenty-first data science in its application to people. And yet as critical data studies scholars have shown, recent deployments of data science risk a bevy of injustices, which these scholars variously conceptualize as "automated inequality," "default discrimination," and "algorithmic oppression."[7]

As one should expect of any history of the present, Galton remains a complicated predecessor for our twenty-first century data politics. His efforts in the 1880s and 1890s anticipated (but by no means fully realized) what I have elsewhere conceptualized as the emergence of full-scale "informational persons" and their "infopolitics" in the 1920s and 1930s.[8] Our data politics today are steeped in nearly one hundred years of infopolitical formats that fasten us to all manner of systems of datafication.[9] Galton's technologies of measure are part of the pre-history of how such a politics of data first emerged in the generation following him.

---

[5] My formulation here is meant to imply that genetics and genomics have always been data-centric sciences; see Koopman 2019 xxx (*Hastings Center Report* article) arguing for this. In the case of early Galtonian genetics, we do not find in his work an anticipation of molecular biology approaches to genetics, but we do find there conceptual models drawing on bioinformatics. These models are arguably more central to genetic sciences than are physico-chemical studies of structure.

[6] Cite Foucault D&P "history of present" xxx.

[7] See Eubanks xxx, Benjamin xxx, and Noble xxx

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

[9] See specifically on formats as techniques of infopolitics Koopman 2021 (*Political Theory* article) and Koopman, *Our Data*, pp. xxx.

xxxCITE PT article forthcoming

In what follows, I focus specifically on a Galtonian legacy that remains with us today. This legacy continues to take pride in the data of measure but has also tried to pay its dues in disowning Galton's regret. This should not sound altogether unfamiliar. It is the self-image operative in many contemporary deployments of data-driven inquiry, including much data science: a deep reliance on measurement data, a use of those data for projects in (among other things) social amelioration, and an innocent disavowal of such odious bigotries as racism.

This contemporary self-image needs be interrogated through the critical distinction between structural racism (sometimes referred to as institutional racism) and attitudinal racism (sometimes simply called racial prejudice).[10] This distinction demands of us that we interrogate our practices not just for overt racist attitudes (of which the contemporary virtuous citizen is presumably not possessed) but also for covert structures that reproduce racialized patterns of inequality. A concept of structural racism, in other words, alerts us to the possibility that ongoing racial inequality is not just an unwanted after-image of centuries of intentional racial domination but is more pressingly an integral part of the practical functioning of our social institutions.

Leveraging this distinction, I here interrogate the possibility that among the many elements of our inegalitarian social structure are information technologies that directly contribute to the reproduction of inequalities. If social structure can be conceived in terms of what the political philosopher John Rawls called "the basic structure of society,"[11] then it is a crucial question for a highly-technologized society such as ours whether technologies are significant components of that basic structure.[12] I here respond to this question by taking it up with respect to the specific issue of how social structure reproduces the distribution of racial inequalities through information technologies. In other words, the question I here raise concerns whether, and how, data technology is a crucial component in the structural maintenance of inequalities of race.

My argument thus challenges the comforting self-image noted above. Restricting ourselves to only the most odious form of racism, that of attitudinal racism, we tell ourselves that we can keep separate what Galton could not. We tell ourselves we are not racist after the model

---

[10] Xxx See Carmichael/Hamilton and Bonilla-Silva
[11] Xxx See Rawls
[12] Xxx Cite Winner as asking this question.

of Galton's despicable attitudes that motivated his programs in eugenics. We think of our pride as unencumbered from prejudice. And with that we take our duty to be discharged. We move forward in confidence that we are deploying our technologies of measure without exacerbating racial inequality. What is stunning about this self-image is simply that we also know for a fact that even where racial prejudice is rejected, there nevertheless persist deep racial inequalities. My aim here is to consider the possibility that one explanation for this is that we have yet to seriously interrogate other of Galton's prides that we have inherited. And thus we remain entangled in numerous inequalities, and thereby unwantingly, and maybe even at times unwittingly, entangled in racial inequality.

Yet we do have alternative precedents who exemplify a quite different data pursuits, namely those committed to data equality. One such precedent was a transatlantic contemporary of Galton's, W.E.B. Du Bois. After furnishing examples of Galton's pride from the 1880s and 1890s over the next few sections of the paper, I turn in the final section to instances of information-driven equality pursued by Du Bois in the 1890s and 1900s. In considering this pairing, it is important that we regard Du Bois's strategies as more than just a counterpoint to Galton's pride. Du Bois's work more importantly forms a challenge to our own ongoing presumptions of innocence about the data-driven inequalities we unwittingly perpetrate. They are a call to us to fulsomely pursue equality in our deployments of data and measure.

In presenting Galton and Du Bois, the primary exhibits I shall focus on are informational apparatus that each crafted for international expositions in 1884 and 1900, respectively. These exhibits add further detail to Shannon Mattern's recent survey of the "spectacles of data" showcased at world's fairs from the late-nineteenth to mid-twentieth centuries.[13] When we witness such spectacular data apparatus, their clean functionality easily dazzles us into the comforting belief that they are neutral. The media-archaeological methods informing my sensibility here help disabuse us of such pretenses, for they invite us to consider not just those past data programs that continue to constitute the present but also those that have for too long been left idle and underutilized.[14]

---

[13] See Shannon Mattern, "The Spectacle of Data: A Century of Fairs, *Fiches*, and Fantasies," in *Theory, Culture, & Society* 37, no. 7-8, 2020: 133-155.
[14] Among the media-archaeological and media-genealogical methodological precedents for my approach here are studies by Cornelia Vismann (xxx), Lisa Gitelman (xxx), and Jeremy Packer, et. al. (xxx). See discussion of this

Galton's Pride

Galton's failures in his many attempts to marry measure to race was his prime regret. But measure ever remained his pride. He sought to apply his gauges to nearly all that he looked upon. And surely the greatest of Galton's passions for what he called "grasp and measure" was that of data concerning the influence of heredity.[15] In this Galton was without doubt indulging in that temper of his age most famously associated with his older half-cousin, the evolutionary naturalist Charles Darwin (Galton was born in 1822, Darwin in 1809). The difference between the two is fine but crucial. Whereas Darwin seemed largely content to establish heredity as a vector of influence on the organism and the species, Galton relentlessly pursued the precise measurement of this influence. It was this pursuit that led Galton out of psychological and anthropological science proper into that erstwhile quasi-science of eugenics. For if one could measure degrees of hereditary influence, then one was in a position to submit those measures to comparison, and to detect thereby presumed hereditary differentiations. If Darwin taught his Victorians that the self is not wholly self-made, then Galton tried to teach them that the self is almost wholly made by others, but specifically by parentage.

"Parentage" was the title of the first chapter of Galton's autobiography, published in 1908, some 27 months before he died. After ten pages spent recounting his parents, grandparents, great-grandparents, uncles, aunts, and even his famous half-cousin (he notes of Charles's father Erasmus Darwin that, "His hereditary influence seems to have been very strong"[16]), Galton comes to his point: "The general result of the foregoing is that I acknowledge the debt to my progenitors of a considerable taste for science, for poetry, and for statistics; also… a rather unusual power of enduring physical fatigue without harmful results."[17] Galton's autobiography neatly performed his theory that family mattered much in the making of the self. And for Galton, all that matters can be measured, so that we can say just how much it matters.

---

methodological approach in my collaborative work in Koopman, Jones, Simon, Showler, and McLevey (xxx "When Data Drive Health," *BioSocieties*, pp. 4-7).

[15] Galton 1884b, 180

xxxGalton, Francis. 1884b. "Measurement of Character" in *Fortnightly Review* 36, 1884: 179-185.

[16] Galton 1908, 7

xxxGalton, Francis. 1908. *Memories of My Life*, third edition. London: Methuen and Co., 1908.

[17] Galton 1908, 11

Before considering in more detail below Galton's vision for the measure of family, I turn first to exhibiting some of the more general features of his technologies of measure. These technologies remain with us in more ways than we commonly think. So too does his focus on the family as the crucial part of our inheritance. What I hope to excavate by considering each in turn is the way in which a series of technical and discursive projects intersected in an informational infrastructure capable of producing pursuits of inequality that remain resilient more than a century later.

Measuring Technology

Galton's technologies of measure constitute a technical apparatus that remains very much at the heart of contemporary efforts in fields such as data science. Today's data science computations are run on microprocessors, but their operations remain informed by Galton.

Consider one of Galton's small but deep contributions to statistical reasoning, namely the concept of percentile ranking (he called them "centiles").[18] Today it is routine for us to ask which percentile we fall into amongst a population measured on a standard scale (be it an IQ test or a GRE score). These percentile rankings mean something to us almost automatically. Simple as the idea is, it is easy to forget that somebody had to come up with this method of classing scores. As Galton himself said: "All this is an old tale now, but I had to take a great deal of trouble before it was clearly thought out and well tested."[19] That it was Galton who took the trouble to do so is quite telling of his love of distinction. Galton was less interested in how high (or low) a person scored on a test than in how high (or low) their scores were relative to other test-takers. And ultimately, Galton's interest in statistics had less to do with averages represented in the bulging middle of a curve and more to do with deviations from the mean where the superlative comes into view.[20]

Our inheritance of a statistical apparatus that relies much on such of Galton's technical innovations as percentile rankings, or statistical correlation as later perfected by his admirer and fellow eugenicist Karl Pearson, is a story already well known.[21] Yet many of Galton's smaller

---

[18] See Galton 1908, 268ff. for his own account.
[19] Galton 1908, 268
[20] Porter 1986, 129
[21] Cite xxx ??Stigler 1986 or 2002 or xxx Porter; on Galton and correlation see Hacking 1990, 180ff. and Porter 1986, 128ff..

contributions to sciences of measure are equally poignant. Galton's profound love of measure is perhaps better conveyed by techniques less complicated than any of his profound statistical ingenuities.

I focus here on a second way in which today's data technologies rely on Galtonian innovations. This concerns a series of technical operations that he was keenly aware of but which many data scientists today routinely neglect. This involves apparatus for the collection and normalization of those data that algorithms will later on be able to process. Alongside Galton's contributions to statistical-algorithmic processing are equally important technologies for what I have theorized elsewhere as formatting.[22]

The core technological apparatus for formatting, or the structuring of data, that is found throughout Galton's many projects is the deceptively-simple one of the printed blank form.[23] A sheet of paper with rules and lines, a handful of pre-printed words describing categories, and a set of clearly-defined blank areas (usually called out by underlining or a series of periods forming a kind of proto-under-line) which one is meant to fill out. And so we, and our various researchers (doctors, teachers, scientists), fill out the forms and we become thereby consistent correlates of the forms. We become persons in forms or in-form-ational persons. In Galton's day none of this had yet to consolidate. Galton's forms are for the most part still rudimentary prototypes of the complex, efficient, and rigorous devices that forms would soon become in the decades just after his death. He was part of the story of this consolidation, even if only a rather early part.

Perhaps one of Galton's most famous deployments of the form technology was at his Anthropometric Laboratory. First set up at London's International Health Exhibition in 1884, the laboratory was a small exhibition in which attendees paying threepence could walk through a 6-foot by 36-foot "long narrow enclosure" and take a series of tests measuring such powers as eyesight, reaction time, span of arms, force of blow, and so on.[24] The exhibition itself closed the following year and Galton moved the Laboratory to the South Kensington Museum where it operated another six years. The mental measures Galton was registering were by then already a

xxxHacking, Ian. 1990. *The Taming of Chance*. Cambridge: Cambridge University Press, 1990.
xxxPorter, Theordore. 1986. *The Rise of Statistical Thinking, 1820-1900*. Princeton: Princeton University Press, 1986.
[22] See Koopman 2021, article in *Political Theory* xxx.
[23] See Gitelman (xxxPaper Knowledge, Ch. 1) on the importance of blank forms.
[24] Galton 1908, 245

well-known facet of German measurement psychology.[25]   But interestingly enough, Galton himself had still devised much of the apparatus used in the tests, including "small whistles with a screw plug for determining the highest audible note."[26]   But the most critical apparatus in operation there was Galton's deceptively-simple blank form. The seemingly simple idea of giving a person a series of tests of bodily and perceptual ability was given a new meaning once experts, scientists, and cranks of all kinds began recording those results on standardized printed forms. As Galton was finally closing up his laboratory he found it "a great consolation" to receive "on the very day that I began to dismantle it, the proof sheets of the register, and other forms in many respects like my own, that are to be used in the laboratory at Dublin."[27]   Galton reported that at the Health Exhibition alone "the number of persons measured in the laboratory… was no less than 9,337, and each of them in 17 different ways."[28]   Each of them left with their own copy of the just-produced record of themselves as recorded on a printed blank form (see Figure 1).

[Fig. 1: Anthropometric Lab Record Blank, p. 219 of Galton 1885]
(figures at end or in separate file)

Galton and his assistants retained a carbon copy of each record. And why? "The data collected at my laboratory have been of service in many ways."[29]   In what ways? Galton calculated correlations among different bodily attributes (between lengths of different limbs, between stature and strength). These calculations led him to some of his important contributions to statistical theory, as noted above. But beyond statistics, Galton developed an interest in the use of these records as a measure and index of who a person is.

At the core of his wide-ranging interests were basic actions with data that are still with us today: format, categorize, archive, and search. These more rudimentary operations form a premise for algorithms, both statistical and otherwise. If Galton is remembered most often in the

---

[25] Cite xxx
[26] Galton 1908, 247
[27] Galton 1892, 32
[28] Galton 1885, 206
xxxGalton, Francis. 1885. "On the Anthropometric Laboratory at the Late International Health Exhibition" in *Journal of the Anthropological Institute* 14, 1885: 205-18.
[29] Galton 1892, 32
xxxGalton, Francis. 1892. "Retrospect of Work Done at my Anthropometric Laboratory at South Kensington" in *Journal of the Anthropological Institute* 21, 1892: 32-5.

history of statistics, he should be remembered equally often for formatting techniques that have long since been black-boxed for every statistician launching SPSS, R, or Excel.

Measures of the Family Man

For Galton, who each of us is must be figured as an index of our heredity. In one among many discussions of what we now call the nature-nurture debate, Galton asserted that the signs of "membership" in "race" are "partly personal, partly ancestral." Yet he immediately rejoined that "we need not trouble ourselves about the personal part."[30] All that matters is heredity. Put simply, Galton took family, or ancestry, or heredity, to be the foundation of all of the most interesting distinctions between persons. Family was for him the primary site and server of inequality.

To understand Galton's focus on the family it is crucial to understand that his idea was not just that families ought to work to improve the natural chances of their offspring, but that big-scale public projects should be brought into being that would systematically improve the chances of the 'best' families. As he put it, "Charity refers to the individual; Statesmanship to the nation; Eugenics cares for both."[31]

Galton recognized how difficult it would be to coordinate any eugenics project on the sizeable scale of the nation. For how would one gather and collate such information on families and races nationally? He laments that, "Owing to absence of data and the want of inquiry of the family antecedents of those who fail and of those who succeed in life, we are much more ignorant than we ought to be of their relative importance."[32] That was Galton in his 1883 *Inquiries Into Human Faculty and Its Development*.

But already in his 1869 *Hereditary Genius* Galton had himself performed a number of such inquiries with respect to the measure of intelligence. There he offered detailed studies of "a large body of fairly eminent men" in a survey of every English Judge from 1660, every Statesman from the time of George III (about one hundred years later), and each Premier from roughly the same period, as well as a hodgepodge of Commanders, Writers, Painters, Musicians, Scholars (the capitalizations of these statuses is Galton's). These studies were intended to yield a

---

[30] Galton 1883, 212
[31] Galton 1908, 322
[32] Galton 1883, 212

picture of the family background, class status, and racial membership of all of these successful persons. The obvious premise of the project, given that Galton proposes it as a measure of hereditary genius, was that "high reputation is a pretty accurate test of high ability."[33] Galton was not measuring genius itself, but rather took reputation as its proxy. With this in hand, he was able to show that genius (meaning reputation) is inherited along family lines in the form of "the existence of a law of distribution of ability in families."[34]

Galton pursued and produced numerous such instruments for a grasping measurement of family. One, a small but deep set of published printed blanks from 1884, remains one of the finest anthropometric paradigms for the practices of informatics that would in later decades grasp and measure persons as subjects of data. Galton's *Record of Family Faculties, Consisting of Tabular Forms and Directions for Entering Data*, *with an Explanatory Preface* was published by Macmillan and Co. of London. The volume was produced in a big but slim format: consisting of nearly 70 sheets measuring 8 inches by 11 inches, with a fly-leaf appended between the title page and page 1 on which is reproduced the advertisement in which Galton offered his £500 in prizes to British subjects who provide him with "the best Extracts from their own Family Records."[35] Galton opens his 13-page introduction with this promise: "This book is designed for those who care to forecast the mental and bodily faculties of their children, and to further the science of heredity."[36] Explaining the potential value of such a science, Galton asserts that, "it is possible to foresee much of the latent capacities of a child in mind and body, of the probabilities of his future health and longevity, and of his tendencies to special forms of disease, by a knowledge of his ancestral tendencies."[37] Yet, he observes, "the advance of the science of heredity is seriously delayed through the want of such data," referring precisely to those data that such of his projects as this book and the anthropometric laboratory aim to solicit, normalize, and store.[38]

---

[33] Galton 1869, 2
xxxGalton, Francis. 1869. *Hereditary Genius: An Inquiry into Its Laws and Consequences*, second edition. London: Macmillan and Co., 1892.
[34] Galton 1869, 309
[35] Galton 1884, Fly-Leaf
xxxGalton, Francis. 1884. *Record of Family Faculties, Consisting of Tabular Forms and Directions for Entering Data*, *with an Explanatory Preface*. London: Macmillan and Co., 1884.
[36] Galton 1884, 1
[37] Galton 1884, 1
[38] Galton 1884, 2; Galton later refers to the need for anthropometric laboratories (such as his own) on p. 9.

Galton's book is essentially a data collection machine consisting of some 50 pages of printed blank forms on which a form-filler will write the names of their ancestors,[39] then children,[40] and then describe in detail each of the known family members in terms of vital information (birthdate, birthplace), bodily measures (height, hair color, "general appearance"), mental measures ("mental powers and energy," eyesight), notes on "character and temperament," and finally medical information regarding ailments, major illnesses, and causes of death (see Figure 3).[41]

[Fig. 3: Blank for "Mother's Father" from family development record, p. 22 of Galton 1884][42]

It was a book of printed blanks in which one would give an account of something presumed important about oneself. It was a series of pre-defined boxes into which one fit oneself, one's spouse, one's children, and one's parents.

Galton's family record book serves as a perfect metaphor for a new kind of person that most citizens of democratic bureaucracies would become over the next fifty years: informational persons. The vehicle that would deliver them there was an analytical science of the data of heredity—here encapsulated in the formatting technology of query forms. In the final paragraph of his introduction Galton writes: "The scientific importance of each investigation will, however, be soon appreciated by the author of it, for his researches will lay bare many far-reaching biological bonds that tie his family into a connected whole, whose existence was previously little suspected."[43] These bonds in data, he asserts, give rise to "the conviction that no man stands on an isolated basis, but that he is a prolongation of his ancestry in no metaphorical sense."[44] One turns the page and the truth of it stares right out. Not in a poetic metaphor, but in a table carefully drawn-up and printed blank for the reader to fill in, one is and becomes an "Index to Ancestors."[45]

---

[39] Galton 1884, 14
[40] Galton 1884, 15
[41] Galton 1884, 16-60
[42] In Galton's case the mother's father would have been Erasmus Darwin, "grandfather to me by his second wife; and to Charles R. Darwin (1809-1882), the great naturalist, by his first wife" (1908, 7).
[43] Galton 1884, 13
[44] Galton 1884, 13
[45] Galton 1884, 14

Information-Technological Structural Racism

At the core of so many of Galton's projects was progressive hereditarianism. Hereditarianism is the idea that natural talents, traits, and features (such as intelligence) are primarily inherited—nature not nurture. This is an idea that goes against the grain of the democratic egalitarianism that was beginning to find many footholds in Galton's day. Galton recognized the conflict. "It is in the most unqualified manner that I object to pretensions of natural equality," he once wrote.[46] Such hereditarianism seems to invite a bleak determinism in which we have little role in to play in the dramas of our own lives—one cannot nurture intelligence in oneself but must simply wait for nature to unfurl. But Galton cultivated a progressive branch of hereditarianism that countered this tragic conclusion with a program of uplift. If we cannot nurture the individual, perhaps we can nurture nature itself. Thus was conceived eugenics—good breeding for heritable traits.[47] The term offends nearly everyone today. Yet the idea itself under proximate headings continues to compel many, as Rabinow observed in his essay on Galton.[48] The recent tumult over the upstart science of behavioral genetics is only the clearest example in our contemporary moment.[49]

But even after eugenics has been appropriately denounced, there remains an extremely complicated morality at play in our inheritance of Galton's other pursuits of inequality in which race breeding was not the explicit and leading theme. For those are the pursuits which, like so many of our own today, cannot be simply dismissed for bearing the immoralities of attitudinal racism. These other pursuits, both in Galton and in ourselves, we do not disclaim so readily. Indeed we often altogether miss their moral and political complexity.

Coming to terms with these more complicating configurations is crucial insofar as contemporary data science is today's inheritor of a sizable algorithmic apparatus and formatting technology that Galton in his day helped to perfect. Today's high-performance computing over super-scaled data using machine-learning methods may seem a far cry from Galton's pencil-and-paper statistics and his long-form printed blanks. But both of those Galtonian technologies are

---

[46] Galton 1869, 12

[47] In Galton's family records project, this logical next step for hereditarian inegalitarianism is already made plain in his treatment of certain data fields (namely "birthplace" and "residence") on his forms as proxies or "partial indications" of race (Galton 1884, 5).

[48] Rabinow, "Galton's Regret"

[49] Xxx cite Gideon Lewis-Krause profile of Harden in *The New Yorker*; for another example from the familial (where else?) domain of assisted reproduction see work by my colleague Camisha Russell, *The Assisted Reproduction of Race*, pp. Xxx.

conceptual infrastructures upon which contemporary informational storage and processing rely. Without Galton (or someone else having done exactly what he did) there would be no data science today.

By archaeologically excavating such Galtonian layers of our technological present, we can better see how, despite the widespread admonition of Galton's racist attitudes among today's technological elites, certain of his information technologies are nevertheless among the social structures scaffolding contemporary racial inequalities. Returning to our own contemporary sciences of grasp and measure in light of the more complicated history I have been recounting, I hope to occasion a fuller pause about the fact that we find our contemporary moment saturated in case after case of what Ruha Benjamin calls "default discrimination" and what Virginia Eubanks calls "automated inequality."[50] If these accounts are right, we who are ensconced in data are under an obligation (by our own anti-racist lights) to actively pursue equality in our design of our data technologies.

Benjamin describes how algorithmic decision-making, for example that facilitated by predictive policing algorithms, is discriminatory not by intention but by design. This is because these algorithms are programmed in a way that "builds upon already existing forms of racial domination."[51] For example, even those who are most committed to fairness in these technologies often "still use the crime rate as the default measure of whether an algorithm is predicting fairly, when that very measure is a byproduct of ongoing regimes of selective policing."[52] The result can only be the deepening of entrenched disparities, which in the context of the measure of crime in the U.S. are chiefly disparities of unequal treatment by race. Benjamin further details how even do-good high-tech projects aimed at "technological benevolence" often serve to reproduce already-embedded forms of discrimination despite their intention to counter it.[53]

Eubanks focuses her analysis on projects located at the intersection of social science and social welfare. With a particular eye to family services work, Eubanks details the construction of what she calls "the digital poorhouse."[54] An exemplar is the Allegheny Family Screening Tool

---

[50] Benjamin 2019 and Eubanks 2017
[51] Benjamin 2019, 81
[52] Benjamin 2019, 82
[53] Benjamin 2019, 137-159
[54] Eubanks 2017, 12

(AFST) algorithms employed by a social services agency in rust belt Pennsylvania. The directors of the agency, she reports, "see little downside to data collection because they understand the agency's role as primarily supportive, not punitive."[55] Theirs is a project of state-sponsored uplift. Yet similar to predictive policing algorithms, these systems are beset by the logic of self-fulfilling prophecies. In this case, "a family scored as high risk by the AFST will undergo more scrutiny than other families."[56] Higher scrutiny generates more data and more opportunities for data-based alarms. The intention might be support, but the design promises to mete out, and unequally, punishments.

Just as the statistical and technical infrastructure upon which data science relies stretches far back into the past, the automated discriminations that all this computation has produced are by no means a twenty-first century invention that can be cheerily chalked up to being a beta-version mistake of nascent science. Our data-driven injustices have a longer and deeper history than we would like to believe. That longer history is of interest, because in it we can witness not only early glimmers of the present, but also the fractures and contests in light of which alternatives were, at least at one point, abound.

Du Bois's Pursuit

In the earlier half of the nineteenth century Alexis de Tocqueville famously accused democracy, as represented by America, of a limitless love of egalitarianism: "for equality their passion is ardent, insatiable, incessant, invincible".[57] A generation later, a rough contemporary of Galton's was both one of his age's greatest witness to America's unrealized passions for equality and also one if its greatest innovators of methods in the informational analysis of social inequalities. W.E.B. Du Bois is still today widely affirmed as one of our most powerful progenitors of racial equality.[58] Late in life he wrote of his own "personal life crusade to prove

---

[55] Eubanks 2017, 165
[56] Eubanks 2017, 169
[57] Tocqueville 1835, p. 97 (Second Book, Chapter I). New York: Vintage Books, 1990.
[58] Given the widely divergent interpretations of Du Bois's work in the literature, I note that my presentation of the early Du Bois is one I take to be generally in line of Michael Dawson's characterization of him as a liberal radical egalitarian and Charles Mill's presentation of him as a black radical liberal. See Michael Dawson, *Black Visions: The Roots of Contemporary African-American Political Ideologies* (Chicago: University of Chicago Press, 15; and Charles Mills, "W.E.B. Du Bois: Black Radical Liberal" in N. Bromell (ed.), *A Political Companion to W. E. B. Du Bois*, University Press of Kentucky, 2018, 19-56.

Negro equality."[59] As part of that crusade, Du Bois was also a pioneering sociologist of the late-nineteenth and early-twentieth century whose contributions to both quantitative and qualitative analysis long went unrecognized.[60]

Du Bois's early innovations in social science from the 1890s and 1900s were roughly contemporary with Galton's. Both men preceded the onrushing wave of datafication that would crash over democratic nations in the 1920s and 1930s. Each was equally prescient in sighting what the tides would soon bring. But where Galton's prescience offers us instruction in how technical innovations can be harnessed by dreams of inequality, Du Bois shows how the same can be put to work in pursuit of equality. Yet it is crucial that Du Bois's contribution not be taken as comforting the data scientist with the thought that their work could be used for good as much as for ill. Such comfort indulges the false promise that data technologies can be neutral in the context of their deployment. Du Bois knew well that they could not be, and long before almost anyone else was even asking the question.

What is most compelling in Du Bois's data designs is that he not only pursued equality *with* data but that he also sought equality *within* the data themselves. Du Bois recognized racializing data as themselves a terrain upon which equality needed pursuing. His data work relied upon, and also amplifies, an understanding that unless one explicitly and fervently pursues equality within the very parameters of their datafication, then inequality is almost surely bound to follow.

I turn now to some of the most forceful exemplifications of equality within data that I find in Du Bois's early publications. I recounted above Galton's employments of his anthropometric forms at the 1884 International Health Exhibition in London. Just sixteen years later, Du Bois was across the channel in Paris at the 1900 Exposition Universelle, where he curated the American Negro Exhibit in the Pavilion of Social Economy. Du Bois's presentation has recently been collected and reprinted in full color for the first time by Whitney Battle-Baptiste and Britt Rusert in their *W.E.B. Du Bois's Data Portraits*.[61] The editors confirm my

---

[59] W.E.B. Du Bois, "The Early Beginnings of the Pan-African Movement," 20 June 1958, W.E.B. Du Bois papers (MS 312), as cited by Aldon Morris, "The Sociology of W.E.B. Du Bois as a Weapon of Racial Equality", *Quaderni di Sociologia*, 83- LXIV, 2020, 11-24, https://doi.org/10.4000/qds.4034.

[60] On Du Bois's early training in what would later be understood as quantitative sociology, see Morris, 20-21; the crucial quantitative dimensions of Du Bois's research program are still too often overlooked by qualitative and humanistic scholars considering Du Bois.

[61] Xxx Cite (New York: Princeton Architectural Press, 2018).

claim above that "the Du Bois data portraits reflect a moment just before the disciplines had hardened" such that we can witness in them the work of "looking forward to the forms of data collection and representation that would become central to representations and surveys" over the next century.[62]

What is most remarkable about the data portraits is their unabashed representation of positive facts of African-American progress and equality. In the words of Du Bois's collaborator on the exposition, Thomas Calloway, the exhibit will display "the Negro's development in his churches his schools, his homes, his farms, his stores, his professions and pursuits in general."[63] In Du Bois's own words, the exhibit offered "a series of striking models of the progress of the colored people, beginning with the homeless freedman and ending with the modern brick schoolhouse and its teachers."[64] By featuring African-American "development" and "progress", Du Bois and his collaborators skillfully countered the discourse of the hereditarians, many of whom would have been present in Paris with their own exhibits (though Galton himself seems not to have been).[65]

Consider Plate 47, titled "Illiteracy of the American Negroes compared with that of other nations." This is a classic bar chart with ten measures protruding from left to right, each bar labeled (in French). No exact percentages are given, and so the chart serves a purely comparative purpose. At the top of the chart are, "Roumanie", "Servie", and "Russie" each indicating roughly the same level of illiteracy. Just below them, showing significantly less illiteracy, is a bar labeled "Negroes, U.S.A." followed by "Hongrie" with only slightly less illiteracy. The last five bars are all Central and Western European nations with comparatively lower illiteracy, though that in Italy is not much lower than in Hungary, and Sweden (at the bottom of the chart) is shown to have remarkably higher literacy than even France (second from last).

[Fig. 4: Du Bois, Plate 47, from Battle-Baptiste and Rusert 2018]

---

[62] Battle-Baptiste and Russert, *Data Portraits*, 13; see also Morris, *Scholar Denied*.

[63] Thomas Calloway, "The Negro Exhibit," in *Report of the Commissioner-General for the United States to the International Universal Exposition, Paris, 1900,* vol. II (Washington: GPO, 1901), 463-467 (get exact pg xxx).

[64] W.E.B. Du Bois, "The American Negro at Paris," *American Review of Reviews* (Nov. 1900), 576 (full pages 575-577).

[65] For a criticism of Galton, much later in life, see W.E.B. Du Bois, "Equality of the Races," ca. 1955, in W. E. B. Du Bois Papers (MS 312), Special Collections and University Archives, University of Massachusetts Amherst Libraries, https://credo.library.umass.edu/view/full/mums312-b207-i018, p. 3 (labeled '14' in top-left corner).

In describing the exhibition the following year, Du Bois explicitly called attention to this particular data portrait.[66] The editors of the reprint note that the chart serves to unambiguously "correct misconceptions about the education of black Americans."[67] For it visually demonstrates that inequalities in education owe more to socio-historical factors than biological-racial causes.

Du Bois's data portraits in Paris are not the only instance of his egalitarian science of data. Equally striking is his 1899 study *The Philadephia Negro*, produced on the back of a gargantuan task of compiling a huge volume of data on an urban black population, the first study of its kind on any demographic in any city.[68] This book offers insight into Du Bois's data collection methods including those that may have been employed for the Paris exhibition the next year. Of particular interest are its appendix reproductions of the questionnaires or "schedules" used to generate study data.[69] Du Bois's schedules are functionally quite like Galton's anthropometric records discussed above—their very construction is designed to elicit data inputs. But where Galton's forms belie formats in search of familial inequalities, Du Bois's formats are a formidable alternative in their function of tuning inquiry to the many modes of development possible within a politically-neglected population. One example of this that would resonate in Du Bois's presentations in Paris the next year are his studies of the growth of literacy rates among black Philadelphians.[70] He even included in one footnote a table with a bar graph showing that the literacy rate of a sample black population in Philadelphia was above that in five European nations, and only slightly below that in Germany.[71] In contrasting Du Bois's schedules to Galton's forms, what stands out is the former's commitment to collecting those data that would reveal not just the influence of an individual's heredity, but also the influences of their social environment. In this, Du Bois's data methods are designed to make space for evidence of those equalities among persons that Galton simply assumed away.

From these and other of Du Bois's pursuits of equality follows a crucial imperative: *where data is used to grasp and measure our social conditions, even when it is for the sake of*

---

[66] Du Bois, "The American Negro at Paris," *Review of Reviews*: 575-577 at 577.
[67] Battle-Baptiste and Rusert, *Data Portraits*, Plate 47
[68] W.E.B. Du Bois, *The Philadelphia Negro: A Social Study* (Oxford: Oxford University Press, 2007 [orig. pub. 1899]).
[69] Du Bois, *Philadelphia*, 276-286
[70] See questions 9 and 10 on the form on Du Bois, *Philadelphia*, 276 as well as the discussion based on these data at Du Bois, *Philadelphia*, 64.
[71] Du Bois, *Philadelphia*, 68n8

their amelioration, our very design of data must be actively and fervently trained on equality, for otherwise it is just too hard to not reproduce inequality. This charge is crucial insofar as data-driven social projects are particularly susceptible to inegalitarianism, a tendency that Du Bois spent a career witnessing time and again. It is as if inequality is the default condition for data-driven social science, at least where the object of social inquiry is a society deeply riven by inequality.

Toward Equality in Data

There is equality and there is equality. My argument is not just that data can and should be used to pursue equality, for this is an argument that is already widely familiar and hardly contested. Rather, my point is that those who do anything with data (including pursuing equality) need to be fervently attentive to ways in which inequalities may be designed into their data. Critical data studies scholars like Benjamin and Eubanks have shown how the use of data in pursuit of equality can go awfully awry. Du Bois helps us correctively understand that the pursuit of equality *within* data is a condition of the pursuit of equality *with* data.

What, then, is the pursuit of equality in data? It involves resolute attentiveness to data formats, including for instance the relevant fields and permissible variables internal to any datafication. It involves unflagging focus on the dangers of innocent-seeming proxy fields for politically-charged social categories. It involves explicit interrogation into whether the measuring instruments employed to make data are themselves reproductive of social conditions they might be charged to ameliorate, as exampled by racial bias in intelligence testing instruments, to take another case in which Du Bois anticipated later critical scholarship.[72] Without pursuing such egalitarianism within data, we leave those whose lives are disclosed by data too much exposed to the haunting hierarchies of manifold legacies of inequality.

The data that are given to us may not yet be ready to beat against the currents of our history. There truly is nothing in the very idea of data that fosters inequality rather than equality. It is our choice whether we design and deploy databases, information architectures, and algorithmic processing apparatus that generate or mitigate inequality. And yet such choices are deeply burdened by the histories in whose futures we remain buried, and therefore also by

---

[72] See W.E.B. Du Bois, "Race Intelligence" (from *The Crisis*, Jul. 1920) in Du Bois, Nathan Huggins, ed., *Writings*, (New York: Library of America, 1987, 1181-1183.

present social and technological contexts within which these choices become operative. While it may seem to some that it is easy to choose for equality, the history of our present teaches us how hard the choice has been to make.

References

xxx

Table of Figures (see attached on following pages)

ANTHROPOMETRIC LABORATORY,

Arranged by Francis Galton, F.R.S.

Sex	Colour of eyes

EYESIGHT.

	right eye	left eye
Greatest distance in inches, of reading "Diamond" type }		

Colour-sense, goodness of }	

JUDGMENT OF EYE.

Error per cent. in dividing a line of 15 inches } in three parts	in two parts

Error in degrees of estimating squareness }	

HEARING.

Keenness can hardly be tested here owing to the noises and echoes.

Highest audible Note }	between	$\left\{ \begin{array}{l} 0\cdot000 \\ \text{and} \\ 0\cdot000 \end{array} \right\}$	vibrations per second.

BREATHING POWER.

Greatest expiration in cubic inches }	

Date	Initials

SWIFTNESS

of blow of hand in feet per second }	

STRENGTH

of squeeze in lbs. of } right hand	left „	of pull in lbs. }

SPAN OF ARMS

From finger tips of opposite hands }	feet,	inches.

HEIGHT

Sitting, measured from seat of chair }	feet,	inches.
Standing in shoes	feet,	inches.
less height of heel...................................		inches.
Height without shoes	feet,	inches.

WEIGHT

in ordinary in-door clothing in lbs. }	

1. Date of birth.   Birthplace.
....

2. Occupation.   Residences.
....

3. Age at   { The place for this entry
   marriage.   { is at 4 in next page.

4. do. of wife   { The place for this entry
                   { is at 3 in next page
....

5. Mode of life so far as
   affecting growth or health.
....

6. Was early life laborious ? why and how ?
....

7. Adult height.   Colour of hair when adult.   Colour of eyes.
....

8. General appearance.
....

9. Bodily strength and
   energy, if much above
   or below the average.
....

10. Keenness or imperfection
    of sight or other senses.
....

11. Mental powers and
    energy, if much above
    or below the average.
....

12. Character and
    temperament.
....

13. Favorite pursuits
    and interests.
    Artistic aptitudes.
....

14. Minor ailments to { in youth.
    which there was
    special liability   { in middle age.
....

15. Graver illnesses   { in youth.
                       { in middle age.
....

16. Cause and date of death, and age at death.
....

17. General remarks.

Illiteracy of the American Negroes compared with that of other nations.

Proportion d' illettrés parmi les Nègres Americains comparée à celle des autres nations.
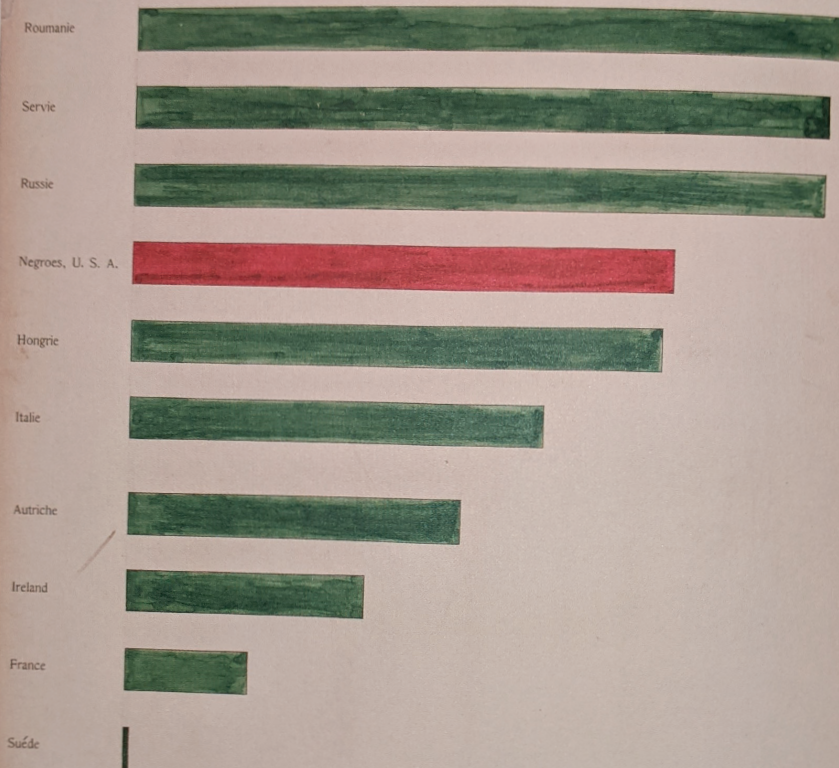
Done by Atlanta University.

Roumanie

Servie

Russie

Negroes, U. S. A.

Hongrie

Italie

Autriche

Ireland

France

Suéde

[47]

Plate 47 A bright red bar set in a horizontal grouping of green bars compares the illiteracy of black Americans to other nations participating in the Paris Exposition. In order to correct misconceptions about the education of black Americans, Du Bois compares the United States to other countries with rates of greater and lesser illiteracy. The key is written in French only, an indicator that Du Bois and his team sought to target the fair's European audience.