

nautil.us

How Eugenics Shaped Statistics

Aubrey Clayton

42–53 minutes

In early 2018, officials at University College London were shocked to learn that meetings organized by “race scientists” and neo-Nazis, called the London Conference on Intelligence, had been held at the college the previous four years.

The existence of the conference was surprising, but the choice of location was not. UCL was an epicenter of the early 20th-century eugenics movement—a precursor to Nazi “racial hygiene” programs—due to its ties to Francis Galton, the father of eugenics, and his intellectual descendants and fellow eugenicists Karl Pearson and Ronald Fisher. In response to protests over the conference, UCL announced this June that it had stripped Galton’s and Pearson’s names from its buildings and classrooms. After similar outcries about eugenics, the Committee of Presidents of Statistical Societies renamed its annual Fisher Lecture, and the Society for the Study of Evolution did the same for its Fisher Prize. In science, these are the equivalents of toppling a Confederate statue and hurling it into the sea.



Dark Prophets: Karl Pearson (left) referred to eugenics as “the directed and self-conscious evolution of the human race,” which he said Francis Galton (right) had understood “with the enthusiasm of a

prophet. ^{Wikipedia}

Unlike tearing down monuments to white supremacy in the American South, purging statistics of the ghosts of its eugenicist past is not a straightforward proposition. In this version, it’s as if Stonewall Jackson developed quantum physics. What we now understand as statistics comes largely from the work of Galton, Pearson, and Fisher, whose names appear in bread-and-butter terms like “Pearson correlation coefficient” and “Fisher information.” In particular, the beleaguered concept of “statistical significance,” for decades the measure of whether empirical research is publication-worthy, can be traced directly to the trio.

Ideally, statisticians would like to divorce these tools from the lives and times of the people who created them. It would be convenient if statistics existed outside of history, but that’s not the case. Statistics, as a lens through which scientists investigate real-world questions, has always been smudged by the fingerprints of the people holding the lens. Statistical thinking and eugenicist thinking are, in fact, deeply intertwined, and many of the theoretical problems with methods like significance testing—first developed to identify racial differences—are remnants of their original purpose, to support eugenics.

It’s no coincidence that the method of significance testing and the reputations of the people who invented it are crumbling simultaneously. Crumbling alongside them is the image of statistics as a perfectly objective discipline, another legacy of the three eugenicists. Galton, Pearson, and Fisher didn’t just add new tools to the toolbox. In service to their sociopolitical agenda, they established the statistician as an authority figure, a numerical referee who is by nature impartial, they claimed, since statistical analysis is just unbiased number-crunching. Even in their own work, though, they revealed how thin the myth of objectivity always was. The various upheavals happening in statistics today—methodological and symbolic—should properly be understood as parts of a larger story, a reinvention of the discipline and a reckoning with its origins. The buildings and lectures are the monuments to eugenics we can see. The less visible ones are embedded in the language, logic, and philosophy of statistics itself.

Francis Galton, Statistical Pioneer, Labels Africans “Palavering Savages”

Eugenics, from the Greek for “well born,” was the brainchild of Galton, a well-born Victorian gentleman scientist from a prominent English family. Thanks to his half-cousin Charles Darwin, Galton adopted the theory of evolution early, with a particular interest in applying it to humans. He posited that characteristics of successful people came from nature rather than nurture (a phrase he coined) and were therefore heritable, so breeding among elites should be encouraged by the state while breeding among those “afflicted by lunacy, feeble-mindedness, habitual criminality, and pauperism” discouraged. By being selective, like choosing the best traits of horses or cattle, he argued we could reshape the human species and create “a galaxy of genius.”

In his 1869 book *Hereditary Genius*, he tallied up famous people in various walks of life who also had famous relatives, to estimate how strongly natural ability could be expected to run in families. But these calculations also moved in predictably racist directions. In a chapter called “The Comparative Worth of Different Races,” he assessed that “the average intellectual standard of the Negro race is some two grades below our own,” which he attributed to heredity. Galton expressed a frequent loathing for Africans, whom he called “lazy, palavering savages” in a letter he wrote to *The Times* advocating the coast of Africa be given to Chinese colonists so that they might “supplant the inferior Negro race.”



Down with Eugenics: Like the removal of Confederate statues from public places in the South, universities and science associations have removed the names of Galton, Pearson, and Fisher from buildings and science prizes.AP

In the world of statistics, Galton is known as the inventor of the

fundamental ideas of regression and correlation, related ways of measuring the degree to which one variable predicts another. He also popularized the concept that the spread of human abilities like intelligence tends to follow a normal distribution, or bell curve (an idea featured most prominently in the 1994 book *The Bell Curve*).

Galton's ranking of the races was based on this distribution, which had only recently come into fashion in applied statistics. Since the early 1800s, it was known theoretically that whenever a large number of small independent increments were combined, the sums would follow a bell curve. If, for example, a crowd of people all stand together and each flip a fair coin to decide to take a step forward or backward, after a large number of coin tosses their positions will be distributed as a bell curve, mostly clustered toward the middle with a few in the extremes. At first, scientists used this in astronomy and geodesy, the study of the shape of the Earth, by assuming their measurement errors arose as the sum of a large number of tiny independent defects and should therefore follow a normal distribution. In the 1840s, Belgian social scientist Adolphe Quetelet discovered the shape in the distributions of people's heights and chest sizes, leading him, with poetic flair, to imagine people as deviations from a common ideal, the “average man.” By smoothing out the errors, one could understand the true nature of this average person, like accurately estimating the position of Jupiter from a few observations.

In science, these are the equivalents of toppling a Confederate statue and hurling it into the sea.

One crucial assumption, though, was that the incremental probabilities were the same for all individuals. If, in our example, half the people use fair coins and the other half use lopsided coins with a 60 percent chance of coming up heads, eventually the group will split into a “bimodal” distribution, with two clusters corresponding to the two means. Instead of one “average man” there might actually be two. Quetelet understood this possibility and took pains to analyze his data in groups that could be considered sufficiently similar.

Quetelet's work with the normal curve had a profound impact on Galton because it provided a scale with which to grade people in every conceivable category. However, when he used the curve himself, Galton predicted that it would always apply to “men of the

same race.” (The idea of grouping only by race had some precedent; in the 1860s, French scientists Louis-Adolphe Bertillon and Gustave Lagneau found what they thought was a bimodal distribution in the heights of soldiers in Doubs, which they took as evidence that the population contained two distinct races.) So, Galton imagined one bell curve for white Europeans, one for Africans, one for Asians, and so on. Comparing curves to each other would show how substantial the racial differences were.

Racist attitudes like Galton’s were not uncommon among the British aristocracy at the heights of colonialism, but Galton gave them scientific backing. He had the authority of a world traveler, in the tradition of Victorian naturalists like his half-cousin aboard the Beagle. His assessment of the supposed inferiority of others to white Britons was an important step toward enshrining these ideas as common knowledge, rationalizing the inestimable colonial violence in Asia, Africa, and the Americas.

Meanwhile, support for eugenics was sparse in Galton’s day. Near the end of his life, Galton delivered the lecture “Probability, the Foundation of Eugenics.” Lamenting that the public had not yet come around on eugenics—in particular that people still married “almost anybody” without regard to their potential for superior breeding—he predicted public opinion would be swayed “when a sufficiency of evidence shall have been collected to make the truths on which it rests plain to all.” When that happened, Galton foresaw a revolution, saying, “then, and not till then, will be a fit moment to declare a ‘Jehad,’ or Holy War against customs and prejudices that impair the physical and moral qualities of our race.”

Karl Pearson, Mathematical Pioneer, Praises America’s Slaughter of “the Red Man”

The eugenics movement’s greatest holy warrior was Karl Pearson, the person primarily recognized today as having created the discipline of mathematical statistics. Pearson was an intellectually driven and prolific scholar of many subjects. After graduating from Cambridge he studied physics, philosophy, law, literature, history, and political science before becoming a professor of applied mathematics at UCL. There he became exposed to Galton’s ideas, and the two men had a fruitful collaboration for years. Pearson referred to eugenics as “the directed and self-conscious evolution of the human race,” which he said Galton had understood “with the

enthusiasm of a prophet.”

Pearson had extreme, racist political views, and eugenics provided a language to argue for those positions. In 1900, he gave an address called “National Life from the Standpoint of Science,” in which he said, “My view—and I think it may be called the scientific view of a nation—is that of an organized whole, kept up to a high pitch of internal efficiency by insuring that its numbers are substantially recruited from the better stocks ... and kept up to a high pitch of external efficiency by contest, chiefly by way of war with inferior races.” According to Pearson, conflict between races was inevitable and desirable because it helped weed out the bad stock. As he put it, “History shows me one way, and one way only, in which a high state of civilization has been produced, namely the struggle of race with race, and the survival of the physically and mentally fitter race.”

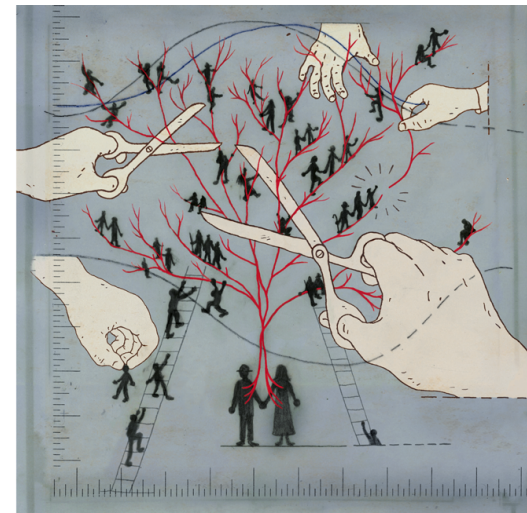


Illustration by Jonathon Rosen

Pearson considered the colonial genocide in America to be a great triumph because “in place of the red man, contributing practically nothing to the work and thought of the world, we have a great nation, mistress of many arts, and able ... to contribute much to the common stock of civilized man.” With an awareness some would criticize this as inhumane, he wrote in *The Grammar of Science*, “It

is a false view of human solidarity, a weak humanitarianism, not a true humanism, which regrets that a capable and stalwart race of white men should replace a dark-skinned tribe which can neither utilize its land for the full benefit of mankind, nor contribute its quota to the common stock of human knowledge."

Pearson the statistician had mathematical skills that Galton lacked, and he added a great deal of theoretical rigor to the field of statistics. In 1901, along with Galton and biologist Raphael Weldon, Pearson founded *Biometrika*, for decades the premiere publication for statistical theory (still highly esteemed today), with Pearson serving as editor until his death in 1936.

One of the first theoretical problems Pearson attempted to solve concerned the bimodal distributions that Quetelet and Galton had worried about, leading to the original examples of significance testing. Toward the end of the 19th century, as scientists began collecting more data to better understand the process of evolution, such distributions began to crop up more often. Some particularly unusual measurements of crab shells collected by Weldon inspired Pearson to wonder, exactly how could one decide whether observations were normally distributed?

Before Pearson, the best anyone could do was to assemble the results in a histogram and see whether it looked approximately like a bell curve. Pearson's analysis led him to his now-famous chi-squared test, using a measure called X^2 to represent a "distance" between the empirical results and the theoretical distribution. High values, meaning a lot of deviation, were unlikely to occur by chance if the theory were correct, with probabilities Pearson computed. This formed the basic three-part template of a significance test as we now understand it:

1. Hypothesize some kind of distribution in the data (e.g., "All individuals are of the same species, so their measurements should be normally distributed"). Today this would be referred to as a "null hypothesis," a straw man standing in opposition to a more interesting research claim, such as two populations being materially different in some way.
2. Use a test statistic like Pearson's X^2 to measure how far the actual observations are from that prediction.
3. Decide whether the observed deviation is enough to knock down

the straw man, measured by the probability, now called the "*p*-value," of getting a statistic at least that large by chance. Typically, a *p*-value less than 5 percent is considered reason enough to reject the null hypothesis, with the results deemed "statistically significant." In Pearson's usage, the word "significant" didn't necessarily have connotations of importance or magnitude but was simply the adjective form of "signify," meaning "to be indicative of." That is, experimental results were significant of a hypothesis if they indicated the hypothesis was true to some degree of certainty.

Applying his tests led Pearson to conclude that several datasets like Weldon's crab measurements were not truly normal. Racial differences, however, were his main interest from the beginning. Pearson's statistical work was inseparable from his advocacy for eugenics. One of his first example calculations concerned a set of skull measurements taken from graves of the Reihengräber culture of Southern Germany in the fifth to seventh centuries. Pearson argued that an asymmetry in the distribution of the skulls signified the presence of two races of people. That skull measurements could indicate differences between races, and by extension differences in intelligence or character, was axiomatic to eugenicist thinking. Establishing the differences in a way that appeared scientific was a powerful step toward arguing for racial superiority.

(Are the science and the scientist so easily separable?

Around the same time that Pearson was describing his method for recognizing non-normal data by writing, "The asymmetry may arise from the fact that the units grouped together in the measured material are not really homogeneous," he was also describing his theory of nations by saying, "The nation organized for the struggle must be a homogeneous whole, not a mixture of superior and inferior races." So, the word "homogeneous," linking the statistical statement to the one from eugenics, had a particularly charged meaning for Pearson, with connotations of racial purity. Homogeneity in data and what it indicated about homogeneity of people had unavoidably racist undertones.

In another typical example, in 1904 he published a study in *Biometrika* that, using a technique of his invention called "tetrachoric correlation," reported roughly the same correlation among 4,000 pairs of siblings for inherited traits like eye color as it did for mental qualities like "vivacity," "assertiveness," and

"introspection." He concluded this meant they were all equally hereditary, that we are "literally forced, to the general conclusion that ... we inherit our parents' tempers, our parents' conscientiousness, shyness and ability, even as we inherit their stature, forearm and span." He ended with a sweeping assertion about the failure of British stock to keep pace with that of America and Germany, advising that "intelligence can be aided and be trained, but no training or education can create it. You must breed it, that is the broad result for statecraft which flows from the equality in inheritance of the psychical and the physical characters in man." In other words, he measured two things—how often siblings' bodies were alike and how often their personalities were—and, finding these measurements to be approximately equal, concluded the qualities must originate in the same way, from which he jumped to dramatic, eugenicist conclusions.

The same year, Galton created the Eugenics Record Office, later renamed the Galton Laboratory for National Eugenics. While working in Galton's Lab, Pearson founded another journal, *Annals of Eugenics* (known today as the *Annals of Human Genetics*), where he could make the case for eugenics even more explicitly. The first such argument, in Volume I of the new journal in 1925, concerned the influx of Jewish immigrants into the United Kingdom fleeing pogroms from eastern Europe. Pearson predicted that if these immigrants kept coming, they would "develop into a parasitic race."

The statistical argument was a one-two punch: by examining a large number of Jewish immigrant children for various physical characteristics, combined with surveys of the conditions of their home lives and intelligence assessments provided by their teachers, Pearson claimed to establish (1) that the children (especially girls) were on average less intelligent than their non-Jewish counterparts, and (2) their intelligence was not significantly correlated with any environmental factor that could be improved, such as health, cleanliness, or nutrition. As Pearson concluded, "We have at present no evidence at all that environment without selection is capable of producing any direct and sensible influence on intelligence; and the argument of the present paper is that into a crowded country only the superior stocks should be allowed entrance, not the inferior stocks in the hope—unjustified by any

statistical inquiry—that they will rise to the average native level by living in a new atmosphere." The journal was immediately referenced by people looking to support their anti-Semitism scientifically. Arthur Henry Lane, author of *The Alien Menace*, said Pearson's conclusions were "of such profound importance as affecting the interests and welfare of our nation, that all men of British race and especially all statesmen and politicians should obtain '*The Annals of Eugenics*.' "



"My Bible": Hitler said he "studied with great interest" the eugenics theories of Galton and his followers. In a letter to Madison Grant, co-founder of the Galton Society of America, Hitler referred to Grant's book *The Passing of the Great Race* as "my bible." Everett Collection / Shutterstock

The most consistent hallmark of someone with an agenda, it seems, is the excessive denial of having one. In introducing the study of the Jewish children, Pearson wrote, "We believe there is no institution more capable of impartial statistical inquiry than the Galton Laboratory. We have no axes to grind, we have no governing body to propitiate by well-advertised discoveries; we are paid by nobody to reach results of a given bias ... We firmly believe that we have no political, no religious and no social prejudices, because we find ourselves abused incidentally by each group and organ in turn. We rejoice in numbers and figures for their own sake and, subject to human fallibility, collect our data—as all scientists must do—to find out the truth that is in them." By slathering it in a thick coating of statistics, Pearson gave eugenics an appearance of mathematical fact that would be hard to refute. Anyone looking to criticize his conclusions would have to first wade through hundreds of pages of

formulas and technical jargon.

When Galton died in 1911, he left the balance of his considerable fortune to UCL to finance a university eugenics department. Pearson, then serving as director of the Galton Lab, was named by Galton as the first Galton Chair in National Eugenics, a position that exists today as the Galton Chair of Genetics. An offshoot of the UCL eugenics department would later become the world's first department of mathematical statistics. In his dual roles as lab director and professor, Pearson had tremendous power over the first crop of British statisticians. Major Greenwood, Pearson's former student, once described him as "among the most influential university teachers of his time."

Most scientists now understand that the data do not speak for themselves and never have.

The work Pearson did during that period was the continuation of Galton's mission to broadcast the "truths" of eugenics in preparation for an overhaul of social norms. This would mean interfering in the intimate relationships within families. Pearson said, "I fear our present economic and social conditions are hardly yet ripe for such a movement; the all-important question of parentage is still largely felt to be solely a matter of family, and not of national importance ... From the standpoint of the nation we want to inculcate a feeling of shame in the parents of a weakling, whether it be mentally or physically unfit."

Given the reactions this was bound to provoke, maintaining the veneer of objectivity was crucial. Pearson claimed he was merely using statistics to reveal fundamental truths about people, as unquestionable as the law of gravity. He instructed his university students, "Social facts are capable of measurement and thus of mathematical treatment, their empire must not be usurped by talk dominating reason, by passion displacing truth, by active ignorance crushing enlightenment." The subtitle of the *Annals of Eugenics* was Darwin's famous quote, "I have no Faith in anything short of actual measurement and the rule of three."

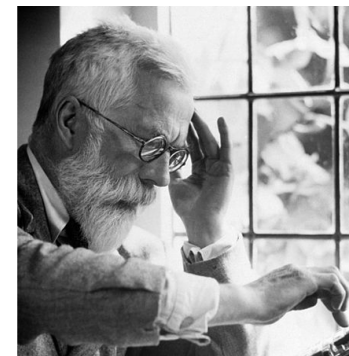
In Pearson's view, it was only by allowing the numbers to tell their own story that we could see these truths for what they were. If anyone objected to Pearson's conclusions, for example that genocide and race wars were instruments of progress, they were

arguing against cold, hard logic and allowing passion to displace truth.

Ronald Fisher, Biological Pioneer, Promotes Sterilization of the "Feeble-Minded"

Ronald Fisher, Pearson's successor both as the Galton Chair of Eugenics at UCL and as editor of the *Annals of Eugenics*, is the only other person with a legitimate claim as the most influential statistician of the 20th century. Fisher was incredibly influential in biology, too. Primarily, his 1930 book *The Genetical Theory of Natural Selection* helped reconcile Mendelian genetics with Darwinian evolution, the project of evolutionary biology called the "modern synthesis." For these and other contributions, he was widely celebrated then and now. In 2011, Richard Dawkins called him "the greatest biologist since Darwin."

But before he was a biologist or a statistician, Fisher was a eugenicist. He became a convert while an undergraduate at Cambridge, where he encountered Galton and Pearson's work and helped form the Cambridge Eugenics Society, with Fisher as student chair of the council. Between 1912 and 1920, he wrote 91 articles for Galton's journal, *The Eugenics Review*. In one of his earliest publications, an essay called "Some Hopes of a Eugenist," he wrote, "The nations whose institutions, laws, traditions and ideals, tend most to the production of better and fitter men and women, will quite naturally and inevitably supplant, first those whose organization tends to breed decadence, and later those who, though naturally healthy, still fail to see the importance of specifically eugenic ideas."



The Gatekeeper: Ronald Fisher's eugenicist proposals were, in some cases, aimed at the statistical academy. In *The Eugenics Review*, Fisher wrote, "A profession must have power to select its own members, rigorously to exclude all inferior types." A. Barrington Brown / Science Photo Library / Wikimedia

This eugenics-flavored nationalism would be a running theme throughout Fisher's career. The final five chapters of *The Genetical Theory of Natural Selection*, comprising about a third of the book, contain a manifesto on the fall of nations, with sections including "The mental and moral qualities determining reproduction," "Economic and biological aspects of class distinctions," and "The decay of ruling classes." Fisher claimed the higher fertility rates among the lower classes would bring down any civilization, including the British Empire, so he proposed a system of limits and disincentives against large families of low social status or immigrants.

By Fisher's time, eugenics had gained momentum as part of larger programs for social reform in the early 20th century. The movement met fierce resistance, particularly from religious institutions, and never really achieved prominence in the U.K. British eugenicists succeeded at turning only some of their ideas into reality, most notably immigration restrictions and one especially awful domestic policy, the Mental Deficiency Act of 1913, which made it so anyone deemed "feeble-minded" or "morally defective" could be involuntarily committed. Because the standards for who qualified were notoriously vague, at one time there were upward of 65,000 people living in such state-operated "colonies." In response to the act, G.K. Chesterton wrote *Eugenics and Other Evils*, in which he derided eugenicists for interfering with people's lives, "as if one had a right to dragoon and enslave one's fellow citizens as a kind of chemical experiment."

Meanwhile, Galton's movement spread to the United States, primarily through the efforts of Harvard professor Charles Davenport, a co-editor of *Biometrika*, who learned about eugenics and its statistical arguments directly from Galton and Pearson. In 1910, Davenport founded the Eugenics Record Office in Cold Spring Harbor, New York, which, like Galton's laboratory, collected data on social and physical traits from several hundred thousand individuals. Applying Galton and Pearson's techniques, Davenport

wrote numerous publications arguing the dangers of interracial marriage and immigration from countries of "inferior" stock. Davenport founded the Galton Society of America, an organization of eugenicists in science with influential connections. They used their positions of power to direct American research in the 1920s and '30s and to lobby, successfully, for measures like marriage prohibitions, restrictions on immigration, and forced sterilization of the mentally ill, physically disabled, or anyone else deemed a drain on society.

To get rid of the stain of eugenics, statistics needs to free itself from the ideal of being perfectly objective.

Today, most people associate eugenics with Nazi Germany, but it was from these American eugenicists and Galton-followers that the Nazis largely took their inspiration. Adolf Hitler once said, "I have studied with great interest the laws of several American states concerning prevention of reproduction by people whose progeny would, in all probability, be of no value or be injurious to the racial stock," and in a fan letter to Madison Grant, co-founder of the Galton Society of America, Hitler referred to Grant's book *The Passing of the Great Race* as "my bible."

Similar sterilization policies would have been illegal in the U.K. at the time, but Fisher and other British eugenicists were working to change that. The eerie similarity to Nazi programs was not coincidental. In 1930, Fisher and other members of the British Eugenics Society formed the Committee for Legalizing Eugenic Sterilization, which produced a propaganda pamphlet arguing the benefits of sterilizing "feeble minded high-grade defectives." Fisher contributed a supporting statistical analysis, based on data collected by American eugenicists demonstrating what they claimed was the degree of heredity for intellectual disabilities.

To strengthen their arguments with additional data, the Society reached out directly to Nazi eugenicist Ernst Rüdin, a chief contributor to the pseudoscientific justifications for the atrocities of Hitler's Germany. Rüdin in turn expressed his admiration for the work of Fisher's Committee. Fisher continued to have disturbingly close ties to Nazi scientists even after the war. He issued public statements to help rehabilitate the image of Otmar Freiherr von Verschuer, a Nazi geneticist and advocate of racial hygiene ideas who had been a mentor to Josef Mengele, who conducted barbaric

experiments on prisoners in Nazi camps. In von Verschuer's defense, Fisher wrote, "I have no doubt also that the [Nazi] Party sincerely wished to benefit the German racial stock, especially by the elimination of manifest defectives, such as those deficient mentally, and I do not doubt that von Verschuer gave, as I should have done, his support to such a movement."

In 1950, in response to the Holocaust, the United Nations body UNESCO issued a statement called "The Race Question" to condemn racism on scientific grounds. Fisher wrote a dissenting opinion, which UNESCO included in a revised version in 1951. He claimed the evidence showed human groups differ profoundly "in their innate capacity for intellectual and emotional development" and concluded that the "practical international problem is that of learning to share the resources of this planet amicably with persons of materially different nature."

As a statistician, Fisher is personally responsible for many of the basic terms that now make up the standard lexicon, such as "parameter estimation," "maximum likelihood," and "sufficient statistic." But the backbone of his contributions was significance testing. Fisher's 1925 textbook *Statistical Methods for Research Workers*, containing statistical recipes for different problems, introduced significance testing to the world of science and became such the industry standard that anyone not following one of his recipes would have difficulty getting published. Among Fisher's most prominent disciples were statistician and economist Harold Hotelling, who in turn influenced two Nobel Prize-winning economists, Kenneth Arrow and Milton Friedman, and George Snedecor, who founded the first academic department of statistics in the U.S.

Fisher promoted significance testing as a general framework for deciding all manner of questions, claiming its logic was "common to all experimentation." Applying the same template from Pearson's chi-squared test to other kinds of problems, Fisher supplied many of the tests we still use today, like Fisher's F-test, ANOVA (analysis of variance), and Fisher's exact test. Many of Fisher's contributions were, therefore, just derivations of the mathematical formulas needed for these tests. In the process, his methods made possible whole new research hypotheses: questions like whether two variables were correlated, or whether multiple populations all had

the same mean.

Like Pearson, Fisher maintained that he was only ever following the numbers where they took him. Significance testing, for Fisher, was a way of communicating statistical findings that was as unassailable as a logical proof. Regarding the method, he wrote, "The feeling induced by a test of significance has an objective basis in that the probability statement on which it is based is a fact communicable to, and verifiable by, other rational minds." As he wrote in 1932, "Conclusions can be drawn from data alone ... if the questions we ask seem to require knowledge prior to these, it is because ... we have been asking somewhat the wrong questions."

Breaking the Legacy of Eugenics and False Objectivity

Today, in response to the protests over Galton, Pearson, and Fisher, defenders have argued that lectures and buildings are named to honor scientific contributions, not people. Statistics professors Joe Guinness of Cornell, Harry Crane of Rutgers, and Ryan Martin of North Carolina State wrote in a comment on the Fisher lecture controversy that "we must recall that the esteem of science is maintained by a collective trust that its achievements are independent of the virtues and vices of the people who achieve them, that recognition is not granted or revoked on the pretense of personal friendship or political positioning, and that we can at once celebrate and benefit from scientific contributions while disagreeing wholeheartedly with the personal beliefs of the scientists responsible for them."

But are the science and the scientist so easily separable?

Without Fisher around to advocate for it, the dominance of significance testing is waning. Last year, a letter signed by over 800 scientists called for an end to the concept of statistical significance, and the leadership of the American Statistical Association issued a blunt decree: "Don't say 'statistically significant.'" The heart of the problem with significance testing is that making binary decisions about homogeneity was never a meaningful statistical task. With enough data, looked at closely enough, some inhomogeneities and statistically significant differences will always emerge. In the real world, data are always signaling something. It just may not be clear what.

In 1968, psychologists David Lykken and Paul Meehl of the

University of Minnesota demonstrated this empirically with an analysis of 57,000 questionnaires filled out by Minnesota high school students. The surveys included a range of questions about the students' families, leisure activities, attitudes toward school, extracurricular organizations, etc. Meehl and Lykken found that, of the 105 possible cross-tabulations of variables, every single association was statistically significant, and 101 (96 percent) had p -values less than 0.000001. So, for example, birth order (oldest, middle, youngest, only child) was significantly associated with religious views, and with family attitudes toward college, interest in cooking, membership in farm youth clubs, occupational plans after school, and so on. But, as Meehl emphasized, these results were not obtained purely by chance: "They are facts about the world, and with $N = 57,000$ they are pretty stable ... Drawing theories from a pot and associating them whimsically with variable pairs would yield an impressive batch of [null hypothesis]-refuting 'confirmations.' " That is, any one of these 105 findings could, according to standard Pearsonian and Fisherian practice, be taken as evidence of inhomogeneity and a significant refutation of a null hypothesis.

Meehl became one of the most outspoken critics of significance testing. In 1978, when the method had by then been the standard for over 50 years, he wrote that "Sir Ronald [Fisher] has befuddled us, mesmerized us, and led us down the primrose path," and that "the almost universal reliance on merely refuting the null hypothesis as the standard method for corroborating substantive theories in the soft areas is a terrible mistake, is basically unsound, poor scientific strategy, and one of the worst things that ever happened in the history of psychology." He described his call to get rid of significance testing as "revolutionary, not reformist."

Similar criticism of the use of statistics in science has a long history. As first argued by psychologist Edwin Boring in 1919, a scientific hypothesis is never just a statistical hypothesis—that two means in the population are different, that two variables are correlated, that a treatment has some nonzero effect—but an attempt at explaining why, by how much, and why it matters. The fact that significance testing ignores this is what economists Deirdre McCloskey and Stephen Ziliak in their 2008 book *The Cult of Statistical Significance* called the "sizeless stare of statistical significance." As they put it, "Statistical significance is not a scientific test. It is a philosophical,

qualitative test. It does not ask how much. It asks 'whether,' "—as in, whether an effect or association simply exists. "Existence, the question of whether, is interesting," they said, "but it is not scientific."

[We cannot escape the fact that statistics is a human enterprise subject to human desire, prejudice, consensus, and interpretation.

Despite the objections, the practice is still very much the norm. When we hear that meeting online is associated with greater happiness than meeting in person, that some food is associated with a decreased risk of cancer, or that such-and-such educational policy led to a statistically significant increase in test scores, and so on, we're only hearing answers to the "question of whether." What we should be asking is what causal mechanism explains the difference, whether it can be applied elsewhere, and how much benefit could be obtained from doing so.

In the context of eugenics, it's understandable why Galton, Pearson, and Fisher would care so much about questions of existence. For the purposes of eugenic discrimination, it was enough to state that distinct racial subgroups existed or there was a "significant" correlation between intelligence and cleanliness or a "significant" difference in criminality, fertility, or disease incidence between people of different socioeconomic classes. The first hypotheses were taxonomic: whether individuals could be considered to be of the same species, or whether people were of the same race. The separation was everything—not how much, what else might explain it, or why it mattered, just that it was there. Significance testing did not spring fully formed from the heads of these men. It was crafted and refined over the years specifically to articulate evolutionary and eugenicist arguments. Galton, Pearson, and Fisher the eugenicists needed a quantitative way to argue for the existence of such differences, and Galton, Pearson, and Fisher the statisticians answered the call with significance testing.

Their attitude that statistical analysis reveals truth without help from the statistician is likewise disintegrating. Most scientists now understand that the data do not speak for themselves and never have. Observations are always possible to interpret in multiple ways, and it's up to the scientist and the larger community to decide which interpretation best fits the facts. Sampling error is not the only kind of error that matters in significance testing. Bias can result from how an experiment is conducted and how outcomes are

measured.

Nathaniel Joselson is a data scientist in healthcare technology, whose experiences studying statistics in Cape Town, South Africa, during protests over a statue of colonial figure Cecil John Rhodes led him to build the website "Meditations on Inclusive Statistics." He argues that statistics is overdue for a "decolonization," to address the eugenicist legacy of Galton, Pearson, and Fisher that he says is still causing damage, most conspicuously in criminal justice and education. "Objectivity is extremely overrated," he told me. "What the future of science needs is a democratization of the analysis process and generation of analysis," and that what scientists need to do most is "hear what people that know about this stuff have been saying for a long time. Just because you haven't measured something doesn't mean that it's not there. Often, you can see it with your eyes, and that's good enough."

To get rid of the stain of eugenics, in addition to repairing the logic of its methods, statistics needs to free itself from the ideal of being perfectly objective. It can start with issues like dismantling its eugenicist monuments and addressing its own diversity problems. Surveys have consistently shown that among U.S. resident students at every level, Black/African-American and Hispanic/Latinx people are severely underrepresented in statistics. But to understand the full depth of what that means about the discipline requires navigating a complex web of attitudes, expectations, symbols, and human emotions. No summary number could do it justice. Daniela Witten of the University of Washington statistics department, the member of the awards committee for the COPSS Fisher Lecture who instigated the movement to change its name, told me, "We need to take an intellectual stance, an academic stance, and a historical stance, but we also need to consider the fact that there are real-life people involved."

Fisher's eugenicist proposals were, in some cases, aimed directly at the statistical academy. In 1917, he wrote in *The Eugenics Review*, "A profession must have power to select its own members, rigorously to exclude all inferior types, who would lower both the standard of living and the level of professional status. In this process the eugenicist sees a desirable type, selected for its valuable qualities, and protected by the exclusive power of its profession in a situation of comparative affluence." Fisher acted as a gatekeeper for

the profession of statistics, encouraging what he considered to be the right types to enter, and as his words make clear, he considered the right type of statistician to be one most like himself. Galton and Pearson, similarly, when they imagined the utopia brought about through eugenics, pictured a society of exclusively Galtons and Pearsons.

Emma Benn is a professor of biostatistics at Mount Sinai and an advocate for changing the name of the Fisher Lecture. When I spoke with Benn, who is African-American, she was quick to point out the controversy over the lecture was about issues larger than Fisher. "Let's reevaluate what we think is a great contribution to the field," she said. "Yes, we can address specifically Fisher, but I hope that that would help us delve deeper into, 'What does it mean in science to belong?' 'How does knowing about these things affect our scientific identity?' When we have this conversation about Fisher, or we have a conversation about race, then all of a sudden it's this intellectual exercise that removes people from having to consider how people like me feel in this field."

Addressing the legacy of eugenics in statistics will require asking many such difficult questions. Pretending to answer them under a veil of objectivity serves to dehumanize our colleagues, in the same way the dehumanizing rhetoric of eugenics facilitated discriminatory practices like forced sterilization and marriage prohibitions. Both rely on distancing oneself from the people affected and thinking of them as "other," to rob them of agency and silence their protests.

How an academic community views itself is a useful test case for how it will view the world. Statistics, steeped as it is in esoteric mathematical terminology, may sometimes appear purely theoretical. But the truth is that statistics is closer to the humanities than it would like to admit. The struggles in the humanities over whose voices are heard and the power dynamics inherent in academic discourse have often been destructive, and progress hard-won. Now that fight may have been brought to the doorstep of statistics.

In the 1972 book *Social Sciences as Sorcery*, Stanislaw Andreski argued that, in their search for objectivity, researchers had settled for a cheap version of it, hiding behind statistical methods as "quantitative camouflage." Instead, we should strive for the moral objectivity we need to simultaneously live in the world and study it.

“The ideal of objectivity,” Andreski wrote, “requires much more than an adherence to the technical rules of verification, or recourse to recondite unemotive terminology: namely, a moral commitment to justice—the will to be fair to people and institutions, to avoid the temptations of wishful and venomous thinking, and the courage to resist threats and enticements.”

Even if we use the most technical language, we cannot escape the fact that statistics is a human enterprise subject to human desire, prejudice, consensus, and interpretation. Andreski challenged us to be honest about factors that influence us and avoid serving unjust masters who would push us toward the conclusions that suit them best. We need the discipline of statistics to be inclusive, not only because it’s the right thing to do and because it enlarges the pool of talented statisticians, but also because it’s the best way to eliminate our collective blind spots. We should try to be objective, not in the impossible sense that Galton, Pearson, and Fisher claimed granted them authority, but in the way they failed to do when they let the interests of the ruling class dictate the outcome of their research before it began.

Aubrey Clayton is a mathematician living in Boston and the author of the forthcoming book Bernoulli’s Fallacy.

Lead image: George Combe / Wikimedia



Get the Nautilus newsletter

Cutting-edge science, unraveled by the very brightest living thinkers.