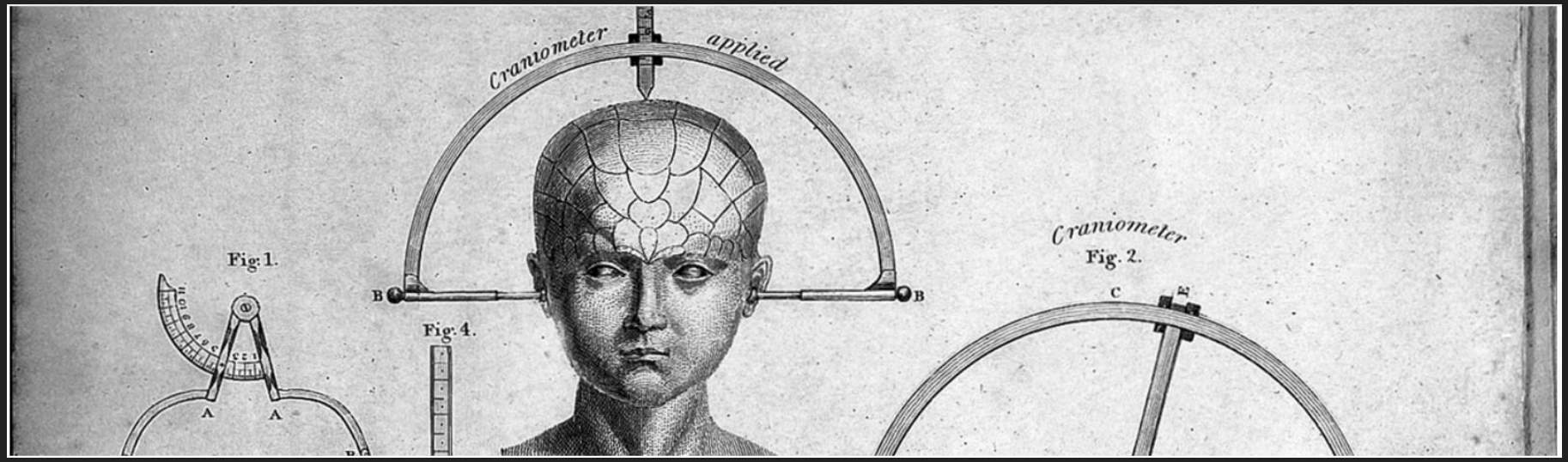
Election day Lecture

The Dark History of Statistics

INFO 6105

November 3, 2020



This year in the US was a season or racial reckoning

Today, we elect a new president

Today, we have a special lecture

Fathers of Eugenics

- 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.”

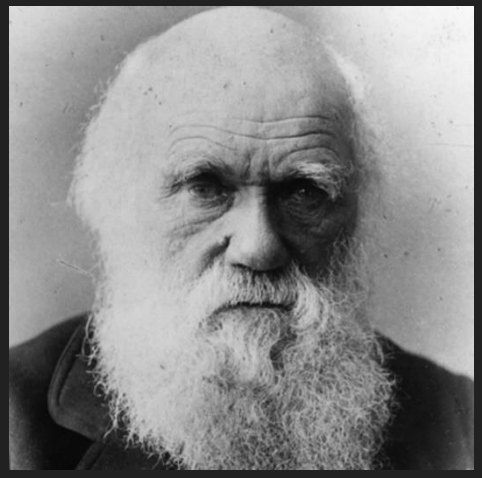
What we now understand as statistics comes largely from the work of Galton, Pearson, and Fisher, whose names appear in terms like *Pearson correlation coefficient, Fisher information, and T-test*

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

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*

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

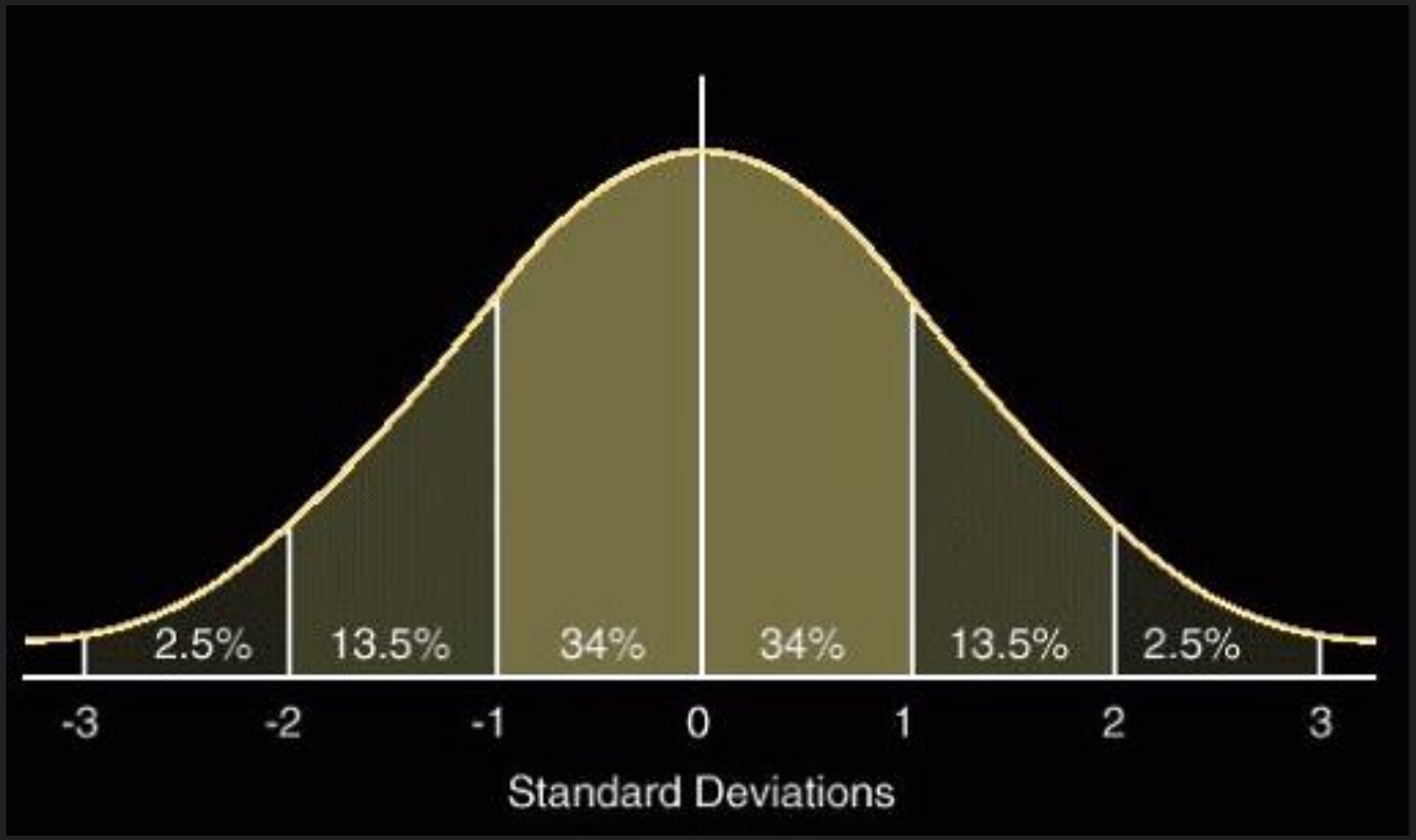
By being selective, like choosing the best traits of horses or cattle, he argued we could reshape the human species and create "*a galaxy of genius*"

In 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*)

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


2.5%
13.5%
34%
34%
13.5%
2.5%
-3
-2
-1
0
1
2
3
Standard Deviations

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

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.


EZRA SHAW/GETTY IMAGES

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

https://www.theguardian.com/news/2017/sep/07/how-the-aristocracy-preserved-their-power

Near the end of his life, Galton delivered the lecture “Probability, the Foundation of Eugenics”

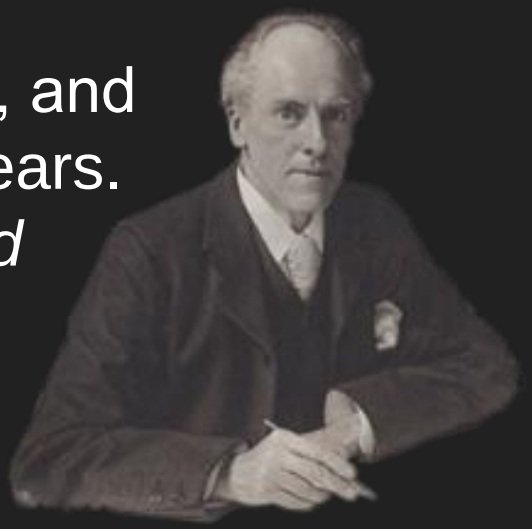
Lamenting that the public had not yet come around on eugenics—in particular that people still married “almost anybody” without regard to their potential for superior breeding—he predicted public opinion would be swayed “when a sufficiency of evidence shall have been collected to make the truths on which 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*”

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

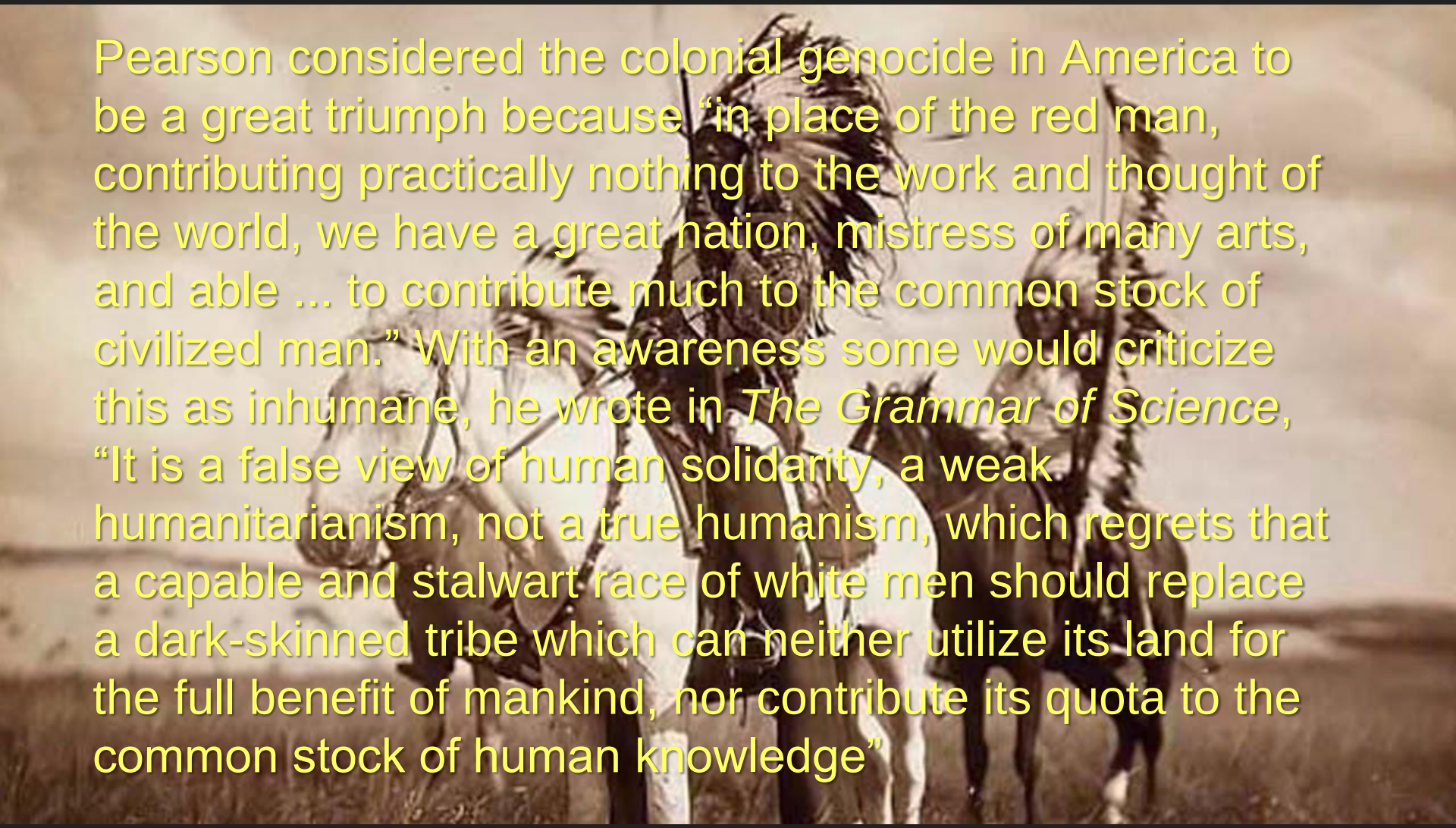
There he became exposed to Galton's ideas, and the two men had a fruitful collaboration for years. Pearson referred to eugenics as "*the directed and self-conscious evolution of the human race*"



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*”



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

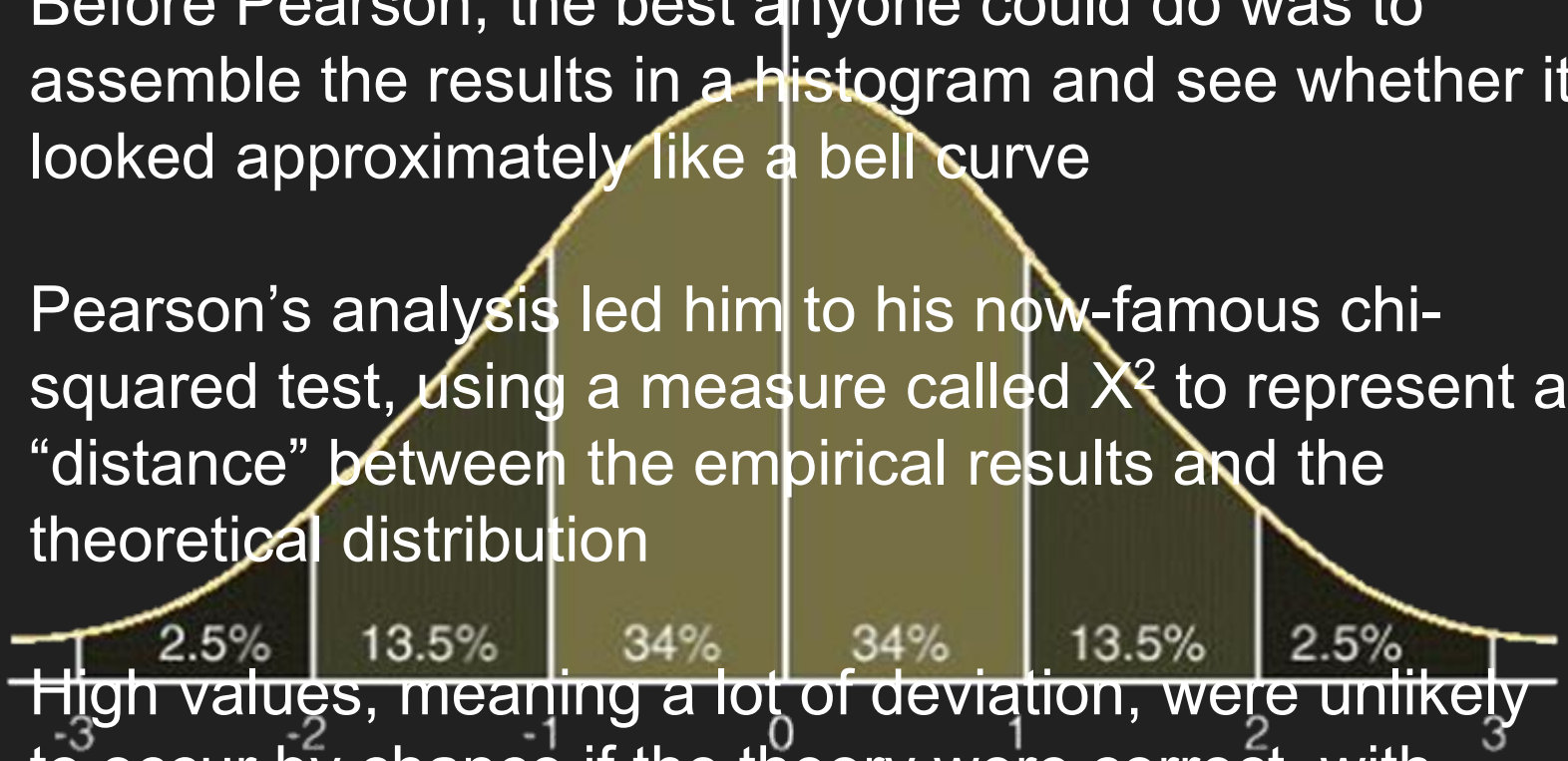
In 1901, along with Galton and biologist Raphael Weldon, Pearson founded *Biometrika*, for decades the premiere publication for statistical theory (still highly esteemed today), with Pearson serving as editor until his death in 1936

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 the *significance test*

Significance testing

- 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.”

- A **t-test tests** a null hypothesis about two means; most often, it **tests** the hypothesis that two means are equal, or that the **difference between** them is zero. ... A **chi**-square **test tests** a null hypothesis about the relationship **between** two variables.

Z-test

- **A null hypothesis, proposes that no significant difference exists in a set of given observations**
- For rejecting a null hypothesis, a test statistic is calculated. This test-statistic is then compared with a critical value and if it is found to be greater than the critical value the hypothesis is rejected
- OR
- A p-value calculates a probability estimate, which we can use to test at any desired level of significance by comparing this probability directly with a significance level

- In a z-test, the sample is assumed to be normally distributed. A z-score is calculated with population parameters such as **“population mean”** and **“population standard deviation”** and **is used to validate a hypothesis that the sample drawn belongs to the same population**
- **Null**: Sample mean is same as the population mean
- **Alternate**: Sample mean is not same as the population mean

T-test, ANOVA, χ^2 test

- **A t-test is used to compare the mean of two given samples**. Like a z-test, a t-test also assumes a normal distribution of the sample. A t-test is used when the population parameters (mean and standard deviation) are not known.
- **Null**: The two samples are the same i.e. sample means are equal
- **Alternate**: The two samples are significantly different
- The statistic for this is called **t-statistic**

- **ANOVA, also known as analysis of variance, is used to compare multiple (three or more) samples with a single test**.
- **Null**: All pairs of samples are same i.e. all sample means are equal
- **Alternate**: At least one pair of samples is significantly different
- The statistic for this is called **F-statistic**

- **Chi-square test is used to compare categorical variables.** There are two type of chi-square test
- 1. **Goodness of fit** test, which determines if a sample matches the population.
- 2. A **chi-square fit test** for two independent variables is used to compare two variables in a contingency table to check if the data fits.
- **Null**: Variable A and Variable B are independent
- **Alternate**: Variable A and Variable B are not independent.
- The statistic for this is called **chi-squared-statistic**
- *Z-test and t-test can be used for data which is non-normally distributed as well if the sample size is large enough* https://www.statisticshowto.com/probability-and-statistics/non-normal-distributions/

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

in 1904, Pearson 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

He confused *correlation* with *cause and effect*

In 1904, 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"

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."

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.”

“*We firmly believe that we have no political, no religious and no social prejudices. 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

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

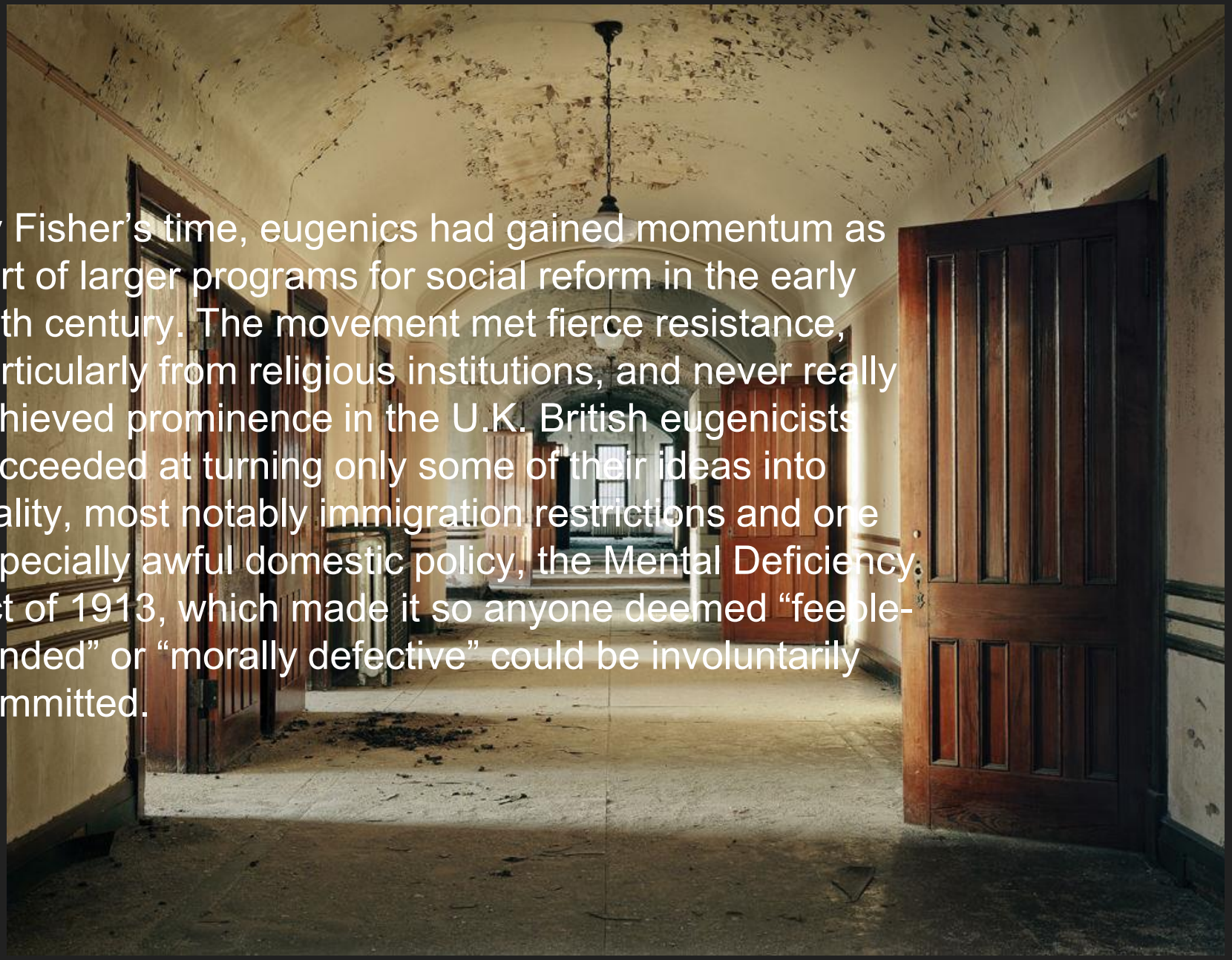
In his dual roles as lab director and professor, Pearson had tremendous power over the first crop of British statisticians. Major Greenwood, Pearson's former student, once described him as "among the most influential university teachers of his time."

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"

In *The Eugenics Review*, Fisher wrote, “A profession must have power to select its own members, rigorously to exclude all inferior types”



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.

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

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.

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.”

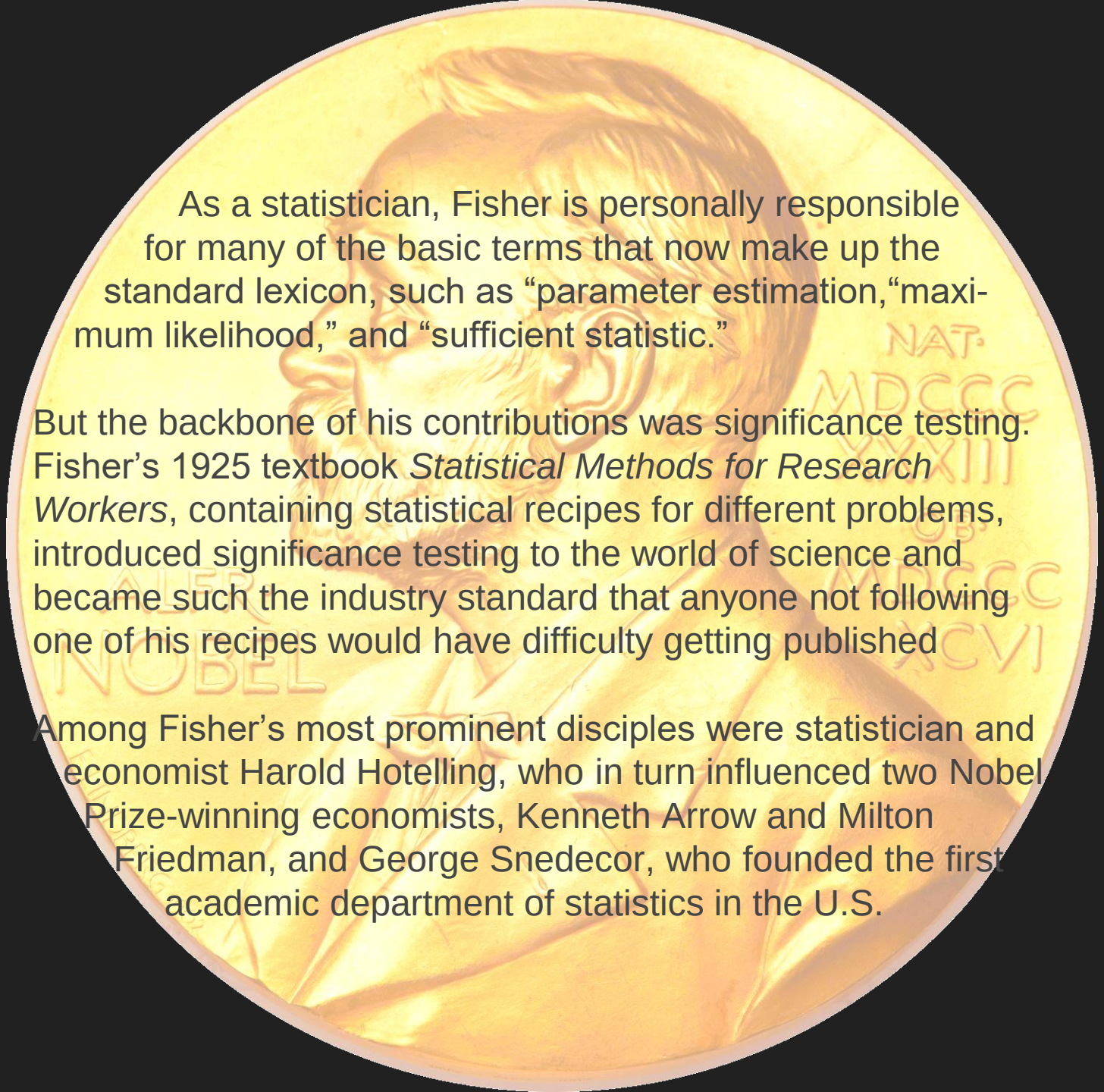
Similar sterilization policies would have been illegal in the U.K. at the time, but Fisher and other British eugenicists were working to change that. The eerie similarity to Nazi programs was not coincidental. In 1930, Fisher and other members of the British Eugenics Society formed the Committee for 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.

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

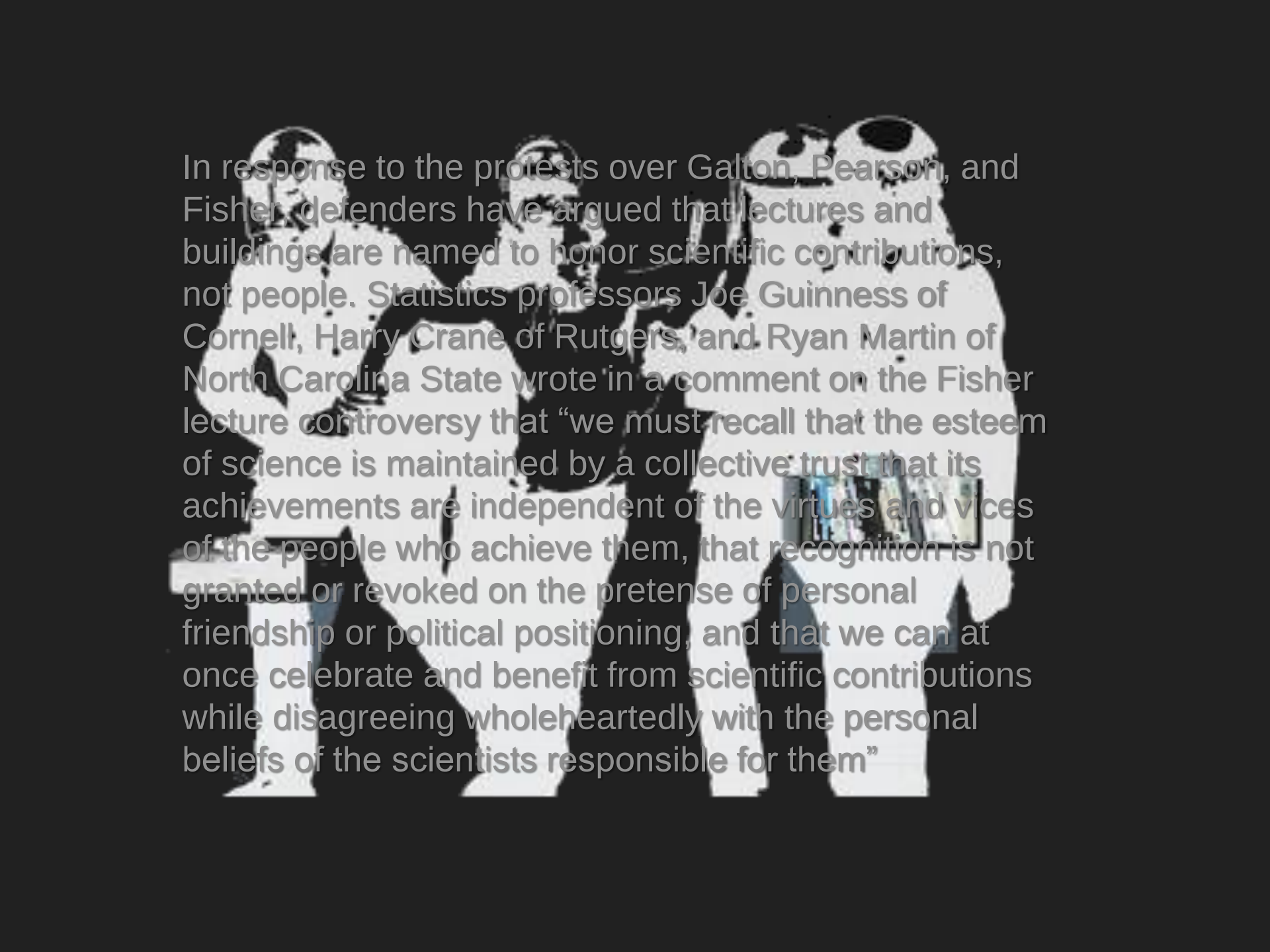
He claimed the evidence showed human groups differ profoundly “in their innate capacity for intellectual and emotional development” and concluded that the “practical international problem is that of learning to share the resources of this planet amicably with persons of materially different nature”



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.



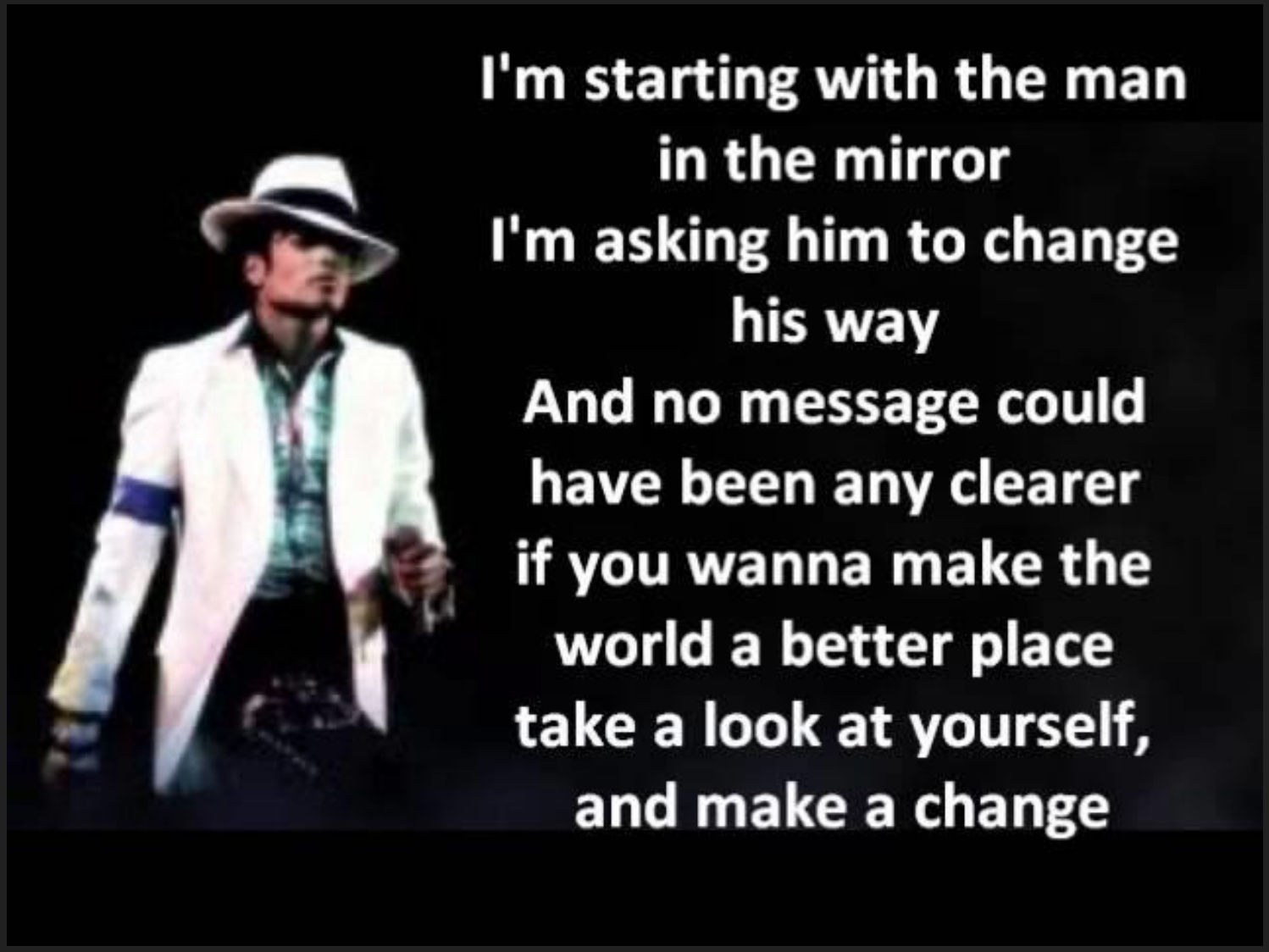
In response to the protests over Galton, Pearson, and Fisher, defenders have argued that lectures and buildings are named to honor scientific contributions, not people. Statistics professors Joe Guinness of Cornell, Harry Crane of Rutgers, and Ryan Martin of North Carolina State wrote in a comment on 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”

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.


I'm starting with the man
in the mirror
I'm asking him to change
his way
And no message could
have been any clearer
if you wanna make the
world a better place
take a look at yourself,
and make a change

Today, India's engineers have thrived in Silicon Valley. So has its caste system.

https://www.washingtonpost.com/technology/2020/10/27/indian-caste-bias-silicon-valley/?utm_source=pocket-newtab

Chinese Han like to advertise celebrations about the cultural influence of its minorities

Can you celebrate diversity in Yunnan and squash it in Xinjian?

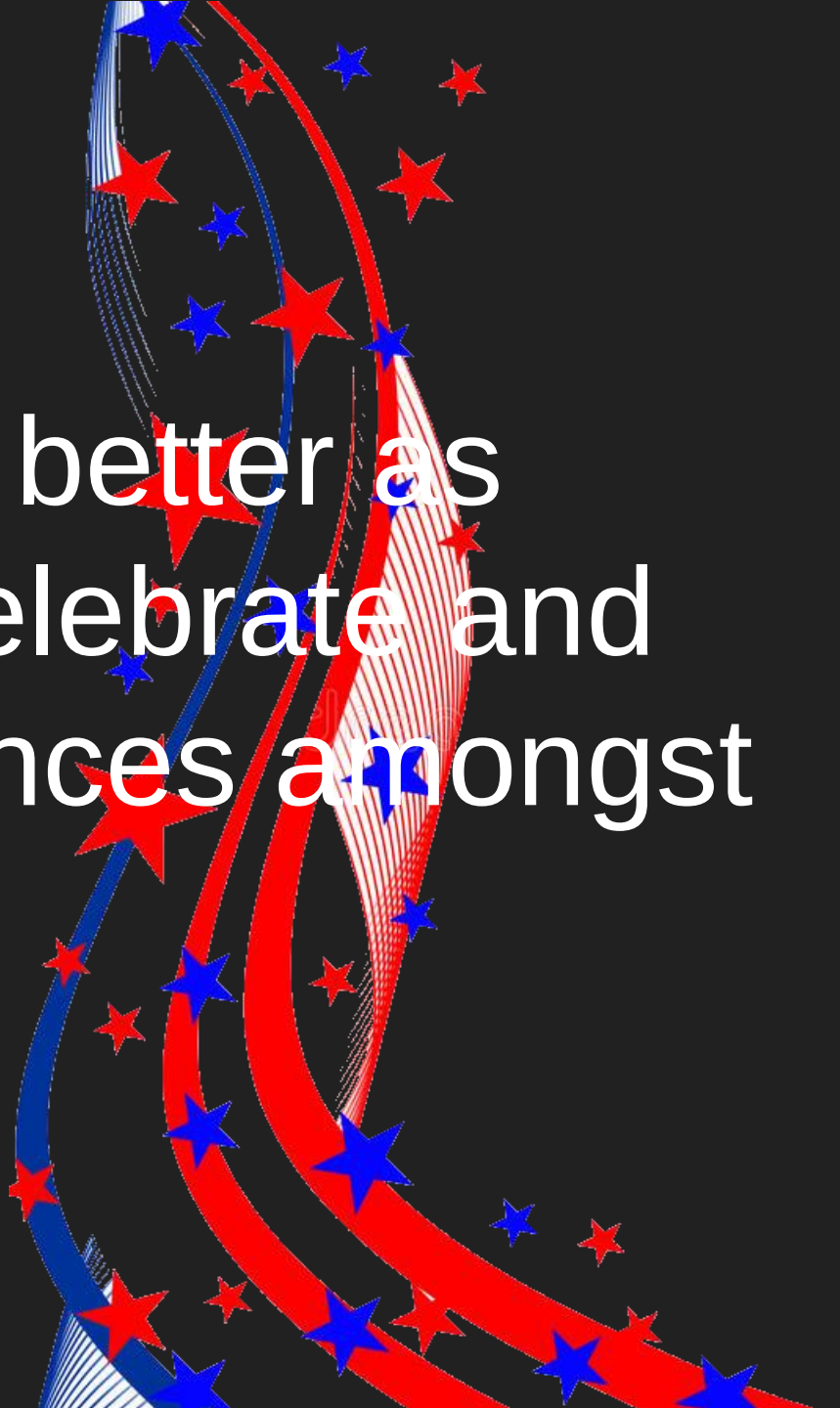
Simulation science
can get rid of bias

Probabilistic programming, like PyMC3, is about using randomness to match models to the data

Genetic algorithms and Machine Learning prove that when a system has only a few traits, it does not have the power to effectively explore all of its state space

In the same way that we celebrate diversity in Nature

We can only get better as societies if we celebrate and value the differences amongst us



They make us *stronger*

Simulations *prove* this

http://nautil.us/issue/92/frontiers/how-eugenics-shaped-statistics?utm_source=pocket-newtab