History of Science Society | THE UNIVERSITY OF CHICAGO PRESS JOURNALS

Francis Galton's Statistical Ideas: The Influence of Eugenics

Author(s): Ruth Schwartz Cowan

Source: *Isis*, Dec., 1972, Vol. 63, No. 4 (Dec., 1972), pp. 509-528

Published by: The University of Chicago Press on behalf of The History of Science Society

Stable URL: https://www.jstor.org/stable/229774

REFERENCES
Linked references are available on JSTOR for this article:
https://www.jstor.org/stable/229774?seq=1&cid=pdf-reference#references_tab_contents
You may need to log in to JSTOR to access the linked references.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at https://about.jstor.org/terms

The History of Science Society and *The University of Chicago Press* are collaborating with JSTOR to digitize, preserve and extend access to *Isis*

JSTOR

This content downloaded from
165.123.34.86 on Thu, 15 Apr 2021 20:07:04 UTC
All use subject to https://about.jstor.org/terms

Francis Galton's Statistical Ideas: The Influence of Eugenics

*By Ruth Schwartz Cowan**

THE PROPER RELATION BETWEEN SCIENCE AND POLITICS is often debated. Some say that the objective pursuit of knowledge is perverted when it is made to serve political ends (witness Lysenko and Aryan physics); others contend that with the world in never-ending crisis a science which is socially irrelevant is luxurious at best and dangerous at worst. In the heat of debate we tend to forget that if the mixture of science and politics is a perversion, it is a perversion which has in the past produced creative scientific work and substantial scientific progress. Science dominated by political ideology is one thing, but science stimulated by political theories and directed toward social goals may be quite another. The development of Francis Galton's statistical ideas is a case in point.

Galton was one of the founders of biostatistics; intellectually and financially he supported the new science in its earliest days. At a time when biology was regarded as a nonquantifiable science he believed that the mathematics of probability could be used to explore biological phenomena, and he lent practical substance to that belief by backing the creation of a new journal, *Biometrika*. In the course of his own researches Galton uncovered two of the most fundamental statistical relationships: *regression*, the relationship between variables which can be measured in the same unit, and *correlation*, the relationship between variables that are not commensurate. The mathematics of correlation and regression have been used in subsequent years in such diverse fields as psychology, genetics, and medicine. This tradition of continuing research began at the Eugenics Laboratory and the Statistics Department of University College, London, institutions which Galton's ample financial resources helped to create.

Francis Galton was a man of many interests; he had an active social life, spent a good part of every year traveling and another good part of his time engaged in various administrative tasks.[1] He was, in a sense, the best kind of dilettante—interested in many things, dabbling in some, pursuing others with great care, but without the discipline of a professional.

The principles of regression and correlation are now easy to state, but in their time they were incredibly difficult to discover. Dilettante that he was, it seems unlikely that

Received July 1971: revised/accepted Apr. 1972.

* History Department, State University of New York at Stony Brook, Stony Brook, New York 11790.

[1] The best source of biographical details about Galton is Karl Pearson, *The Life, Letters and Labours of Francis Galton*, 3 vols. in 4 (Cambridge: Cambridge Univ. Press, 1914–1930). See also Galton's autobiography, *Memories of My Life* (London: Methuen, 1907). Mrs. Galton kept an annual diary which provides information about the Galtons' social life; it is labelled *Louisa's Record* and can be found in the Galton Archive, University College, London.

This content downloaded from
165.123.34.86 on Thu, 15 Apr 2021 20:07:04 UTC
All use subject to https://about.jstor.org/terms

Galton would have suffered the pains of discovery if he had not been strongly motivated; his motivation was, as we shall see, largely political. Galton pursued statistics because he believed that statistics would solve the problem of heredity and that heredity, once understood, could be used to resolve the political and social conflicts that plague the race of men. For Galton this was a real political program, not just empty rhetoric. He sincerely believed that statistics could be used to construct the perfect eugenic state.

In one sense at least, it is not surprising that Galton's most important scientific contributions were made to the science of statistics. He did not possess a sophisticated mathematical mind, but from his earliest days he was an habitual counter and measurer—one might almost say a compulsive counter and measurer.[2] Wherever he went and whatever he did, he usually found something to count or to measure along the way, and often carried this preoccupation to marvelous extremes. While traveling in South Africa, for example, he was overcome by a desire to measure some buxom Hottentot ladies:

> I have dexterously even without the knowledge of the parties concerned, resorted to actual measurement . . . I sat at a distance with my sextant, and as the ladies turned themselves about, as women always do, to be admired, I surveyed them in every way and subsequently measured the distance of the spot where they stood—worked out and tabulated the results at my leisure.[3]

Finding himself bored at a meeting of the Royal Society, Galton constructed a small pocket "ticker" and measured the frequency with which various members of the Society were fidgeting in their seats.[4] Finding himself ill and confined at home he measured the precise temperature of water and pot that would produce a proper cup of tea.[5] Almost every scientific paper that Galton published was concerned, in some manner, with counting and measurement—whether it be means for keeping accurate records of weather conditions or methods for determining the accuracy of geographic instruments or calculations of the frequency of geniuses in given populations or analyses of the degree of resemblance in members of the same family.

Francis Galton habitually mathematized problems that he was studying, whether or not they appeared to lend themselves to mathematical analysis. Thus from the very beginning Galton's study of heredity took mathematical form. His first paper on the subject, "Hereditary Talent and Character," was a statistical analysis of biographical dictionaries, a monument to his predilection for counting.[6] Galton posed the question, "Can extraordinary intellectual gifts be inherited?" and answered the question by

[2] This point was made by Eliot Slater, a psychiatrist, in "Galton's Heritage," The Galton Lecture for 1960, *Eugenics Review*, 1960, *52*: 91–103.

[3] Galton to his brother, Darwin Galton, Feb. 23, 1851, Pearson, *The Life of Galton*, Vol. I, p. 232. For convenience Pearson's book will be cited below as *LLL* with volume and page number directly following ((e.g., *LLL* I 232). Following Pearson's practice, Vol. IV will be indicated as IIIb.

[4] "Measure of Fidget," *Nature*, 1885, *32*:174. Some of the devices that Galton used for surreptitious counting are in the Galton Archive.

[5] *LLL* IIIb 457. The experiments were conducted in Feb. 1859.

[6] Francis Galton, "Hereditary Talent and Character," *Macmillan's Magazine*, June and Aug. 1865, *12*:157–166, 318–327. Francis Galton was Charles Darwin's first cousin, a fact often cited as the most likely reason for Galton's interest in heredity. I am not impressed with the argument. Galton was, by all accounts, a humble man; he may have thought of his eminent cousin as a genius, but he had no such pretension about himself. Heredity could have attracted his attention for several other reasons (about which see below).

This content downloaded from
165.123.34.86 on Thu, 15 Apr 2021 20:07:04 UTC
All use subject to https://about.jstor.org/terms

counting the number of men listed in a biographical dictionary who were relatives of someone else on the list. Finding this number to be greater than the frequency of outstanding men in the general population, Galton concluded that "genius" must be hereditary.

Before 1865, when "Hereditary Talent and Character" was published, Galton had shown no great interest in biology or psychology. He had published articles and books on meteorology and geography, on techniques for traveling in rough areas and techniques for taking accurate geographical measurements, on cyclones and anticyclones—but not on biological problems, and certainly not on the problem of heredity, perhaps the knottiest biological problem of them all.

After 1865 he was as a man transformed; heredity became the overriding interest of his life. Leafing through the pages of "Hereditary Talent and Character" one senses at least part of the reason for this transformation: after reading the *Origin of Species* Galton had become convinced that the study of heredity would yield a viable technique for improving mankind, that over a period of time men could be bred for intelligence and character in precisely the same way that animals are bred for strength or agility.

> The power of man over animal life, in producing whatever varieties of form he pleases, is enormously great. It would seem as though the physical structure of future generations was almost as plastic as clay, under the control of the breeder's will. It is my desire to show . . . that mental qualities are equally under control.[7]

Galton hoped that his ideas about mental heredity would form the basis for a political and moral reformation of society. Once men realized that they could improve future generations by manipulating heredity they would cease searching for environmental solutions to political problems—poor laws, factory laws, sanitation laws—and would begin constructing a society in which eugenic values were supreme. "Hereditary Talent and Character" contains a rudimentary description of such a eugenic utopia, a society which has accepted "race improvement" as the basis of its ethical code. A series of examinations will be given to discover the most talented young men and women in the realm, and they will be honored at public ceremonies and given cash rewards if they choose to marry each other:

> If a twentieth part of the cost and pains were spent in measures for the improvement of the human race that is spent on the improvement of the breed of horses and cattle, what a galaxy of genius might we not create! We might introduce prophets and high priests of civilisation into the world.[8]

Galton realized that this utopia would be impossible without an understanding of the laws of heredity. Since he instinctively believed that natural laws were only comprehensible when mathematized, he spent the rest of his life in pursuit of the mathematics of heredity. His goals were not purely scientific: he wanted to understand heredity so that he could use heredity to elevate men. This was the motive force that drove him toward the discovery of regression and correlation.

The first halting steps on the road to that discovery were taken in 1869 when, with the publication of *Hereditary Genius*, Galton expanded the argument he had made in his earlier paper and introduced the "calculus of probabilities" into his analysis of

[7] *Ibid.*, p. 157.

[8] *Ibid.*, pp. 165–166.

This content downloaded from
165.123.34.86 on Thu, 15 Apr 2021 20:07:04 UTC
All use subject to https://about.jstor.org/terms

mental heredity.[9] Galton knew that in order to make his study truly accurate it would be necessary to determine "ranks" of genius; one would want to know whether the sons of brilliant men were more or less brilliant than their fathers, and in order to know this it would be necessary to have some way of ranking geniuses. Such an evaluation could be achieved, Galton said, by the use of the law of deviations if some standard measurement of genius were available. A large number of men could be given this standard test and their marks on this test could be "curved" (just as examination grades are curved in American universities), making it possible to determine just how extraordinary any given individual was. The normal curve for intelligence could be divided into sections (or ranks) and the number of individuals who could be expected to fall into each section could be computed. In *Hereditary Genius* Galton attempted such a calculation: given a base population of fifteen million (the approximate number of males in Great Britain in 1860) he calculated the number of extremely brilliant and extremely stupid men that could be expected to appear in such a population.[10] Farther than this he was unable to go, lacking any additional data which might be analyzed according to the laws of probability. In the two decades following 1869 a substantial part of his research time was devoted to correcting that lack.

In 1869 Galton had no more than a rudimentary understanding of the laws of probability; essentially he knew no more than what could be found in the pages of Quetelet's *Essai sur l'homme*, first published almost thirty years earlier.[11] Quetelet had pioneered the technique of charting human characteristics on frequency-of-error curves. Using statistics garnered from studies of French and Scottish soldiers he had demonstrated that in several countries, and over several generations, measurements of certain anthropometric characters (such as height and weight) are always distributed in the fashion we now describe as normal. Galton learned about the law of error from William Spottiswoode, the publisher and geologist, an old personal friend:

> My first serious interest in the Gaussian Law of Error was due to the inspiration of William Spottiswoode, who had used it long ago in a Geographical memoir for discussing the probability of the elevation of certain mountain chains being due to a common cause. He explained to me the far-reaching applications of the beautiful law, which I fully apprehended.[12]

Galton was probably familiar with the practical applications of the normal law by virtue of his work in meteorology and instrument testing. In 1858 he had been appointed to the Managing Committee of Kew Observatory, having previously published a few papers on the construction and testing of geographic instruments.[13] At that time Kew functioned as a central station for testing and calibrating instruments; the Gaussian law of error was used repeatedly there (and elsewhere) to determine the

[9] Francis Galton, *Hereditary Genius* (London, 1869), esp. Sec. 1, Chs. 2 and 3.

[10] *Ibid.*, pp. 72–76.

[11] L. A. J. Quetelet (1796–1874) was the Astronomer Royal of Belgium. From the point of view of biostatistics and anthropometry his most significant works were *On Man* (1836; English translation, 1838; republished as *Physique sociale*, 1869), and *Letters on Probability* (1846; English translation, 1849). Although several commentators have assumed that Galton was influenced by Quetelet's work, Pearson disagrees (*LLL* II 12). None of Quetelet's books were in Galton's personal library; his notes on *Physique sociale* fill only one notebook page.

[12] *Memories*, p. 304. The memoir to which Galton referred was William Spottiswoode, "On Typical Mountain Ranges; An Application of the Calculus of Probabilities to Physical Geography," *Journal of the Royal Geographical Society*, 1861, *31*:149–154.

[13] Francis Galton, "Hand Heliostat," *British Association Reports*, 1858, *2*:15–17. On Galton's pocket "altizimuth," see *LLL* II 50.

This content downloaded from
165.123.34.86 on Thu, 15 Apr 2021 20:07:04 UTC
All use subject to https://about.jstor.org/terms

permissible limits of error for any given instrument.[14] Galton may not have realized that the Gaussian law could potentially be used in other sciences, and he may not have realized that it could solve problems quite unlike the determination of errors, but it would have been strange indeed if he had not acquired at least a passing knowledge of the form and function of the law.

Not long after commencing his work at Kew Galton became interested in meteorology, an interest which he retained, actively, almost for the rest of his life.[15] In the 1850s and 1860s meteorologists were addressing themselves to the problem of long-range weather prediction, and they hoped that statistical techniques would prove useful in that regard.

> In treating climatology as a *science*, it is desirable that some correct and convenient mode should be adopted for computing and expressing the comparative *variability* to which the temperature in different parts of the globe, and in different parts of the years in the same place, is subject from non-periodic causes. The *probable variability* computed on the same principle as the *probable error* of each of a number of independent observations has recently been suggested as furnishing an index of the probable daily non-periodic variation.[16]

Galton tackled the problem of reducing and combining large amounts of dissimilar weather data so as to make prediction feasible; he pioneered the use of weather maps, invented a double-drill pantograph for reducing the tracings of self-recording meteorological instruments, and a trace computer which could automatically draw a vapor tension curve on the basis of data from wet- and dry-bulb thermometers.[17]

If one assumes, as Galton did, that every organic character is produced by a number of genetic determinants, then the problem of reducing and combining meteorological data so as to be able to predict the weather in one time at one place is strikingly similar to the problem of predicting the characteristics of an offspring once the constitution of its parents is known. The curve which is produced by Galton's trace computer (a curve which graphs various fixed functions of the ordinates of two other curves) is not unlike the regression curve which Galton was later to discover in the course of his breeding experiments. In 1870 Galton tried to predict wind velocity on the basis of barometric height, temperature, and humidity; he was searching for what statitisticians now refer to as a multiple regression formula.[18] In 1870 he failed to find it, but within a few years a similar formula came to him in the course of studying similar problems

[14] Helen M. Walker, *Studies in the History of Statistical Method* (Baltimore: Williams & Wilkins, 1929), pp. 28–40.

[15] Galton's first meteorological paper appeared in 1861, "Circular Asking for Synchronous Observations. . . ." *Philosophical Magazine*, 1861, *22*:34–35, and his last in 1873, "On the Employment of Meteorological Statistics in Determining the Best Course for a Ship," *Proceedings of the Royal Society*, Apr. 1873, *21*:263–274. In 1868 Galton became a member of the Meteorological Council of the Royal Society, which supervised the Meteorological Office of the Board of Trade; he did not retire from the Council until 1901.

[16] Edward Sabine, "Letter to the Board of Trade," *Proc. Roy. Soc.*, 1854, *7*:342 (italics added). Galton described Sabine as a good friend and a man who "exercised great influence in shaping my future scientific life" (*Memories*, p. 224). The Board of Trade had consulted the Royal Society about the duties that a meteorological office might perform; Sabine, who was chairman of the Kew Committee, responded for the Society.

[17] Francis Galton, *Meteorographica, or Methods of Mapping the Weather* (London, 1863). Drawings of the pantograph and the trace computer are in *LLL* II 45–48. A model of the pantograph is in the South Kensington Science Museum.

[18] Francis Galton, "Barometric Prediction of Weather," *Brit. Assoc. Rep.*, 1870, 31–33; *Nature*, Oct. 20, 1870, N.S. *2*:501–503. Pearson makes the same point, *LLL* II 54.

This content downloaded from
165.123.34.86 on Thu, 15 Apr 2021 20:07:04 UTC
All use subject to https://about.jstor.org/terms

posed by the phenomena of heredity. Here we seem to have a case of intellectual cross-fertilization: Galton (and other meteorologists) suspected that the normal law might be applied with profit to their science; sensing the similarity between meteorological problems and genetical problems he was inspired to transfer the techniques. In this connection it is amusing to note that Galton's moment of inspiration came during a rainstorm:

> . . . the circumstances under which I first clearly grasped the important generalization that the laws of Heredity were solely concerned with deviations expressed in statistical units, are vividly recalled to my memory. It was in the grounds of Naworth Castle, where an invitation had been given to ramble freely. A temporary shower drove me to seek refuge in a reddish recess in the rock by the side of the pathway. There the idea flashed across me, and I forgot everything else for a moment in my great delight.[19]

Thus, in 1869 Galton was no stranger to the law of probability, but he was more familiar with its practical applications than with its mathematical structure. From Quetelet he learned that the normal law might produce information about human populations. Spottiswoode and Kew taught him that it might be used with profit in geography and measurement, while Sabine and the other meteorologists (of whom, coincidentally, Quetelet was one) suggested possibilities for weather prediction. As the years wore on Galton spent more and more time worrying over the uses and implications of probability, but he never became adept at its mathematics; when he had a complex problem to solve he farmed it out to someone else. Although Galton was a compulsive counter and measurer, and although he could handle various *pictorial* representations of the probabilistic laws with ease, he was not a sophisticated mathematician. In a sense this makes his discovery of regression and correlation all the more remarkable.

Hereditary Genius was not well received by the scientific community,[20] but this did not dampen Galton's enthusiasm. He knew that he did not have enough data to apply the normal laws to human populations with any hope of success, and he pondered various means for obtaining the data. In 1873 he suggested that a national registry should be established for developing and administering tests that would determine the physical and mental abilities of the population.[21] Not long after that Galton realized that it might not be necessary to test the entire population in order to determine the distribution of abilities: a random selection of individuals would do. Galton suggested that if 101 members of a population are lined up according to their heights, the tops of their heads would form a curve which we now call a cumulative distribution function. If the height of the twenty-sixth, fifty-first, and seventy-sixth individuals are measured, the mathematics of probability permits the construction of a curve which will describe the distribution of height in the entire population.[22] The task of the national registry would be made that much simpler; 101 individuals could provide information about the whole population.

[19] *Memories*, p. 300.

[20] "Frank's book not well received," Mrs. Galton commented in her diary entry for 1869.

[21] Francis Galton, "Hereditary Improvement," *Fraser's*, Jan. 1873, N.S. *7*:116–130. Galton hoped that each individual who registered would be given a social "rank" and that individuals in the higher ranks would be permitted a greater number of children.

[22] Francis Galton, "Proposed Statistical Scale," *Journal of the Royal Anthropological Institute*, 1874, *4*:136–137; *Nature*, Mar. 5, 1874, N.S. *9*: 342.

This content downloaded from 165.123.34.86 on Thu, 15 Apr 2021 20:07:04 UTC
All use subject to https://about.jstor.org/terms

The public schools also struck Galton as a source of useful data, particularly because they could provide continuous data about changes (if any) in national physique and intelligence. Galton firmly believed that the physique of English school boys was deteriorating as more and more of them were born and raised in cities. According to Galton this was happening because able men are attracted to cities, where the social conditions discouraged early marriages and large families. Intelligent, healthy children are born of intelligent, healthy men. If such men have smaller families than the rest of the population, eventually the offspring of the incompetent and diseased will outnumber the offspring of the intelligent and healthy, and the character of the population at large will have changed—for the worse. Galton disagreed with the social reformers who ascribed deterioration (if any) to the unpleasant living conditions in urban areas. Galton hoped to test his hypothesis on data from the public schools, and he managed over the years to accumulate substantial amounts of data from schoolmasters across the country.[23]

Sometime early in the 1870s Galton turned his attention to the problem of quantifying the phenomena of heredity directly. In 1868 Darwin's theory of pangenesis had appealed to him because it was potentially quantifiable—"the theory of Pangenesis brings all the influences that bear on heredity into a form that is appropriate for the grasp of mathematical analysis"[24]—and in the last chapter of *Hereditary Genius* he had even attempted such an analysis, in a very rudimentary form. His attention was diverted for two years while he undertook, with Darwin's assistance, experimental studies of pangenesis. The experiments failed, much to the dismay of both men, and within a few years Galton began musing once again about mathematization of the laws of heredity.[25] He wrote to Darwin in December 1875,

> It has been an old idea of mine, not yet discarded and not yet worked out, that the number of units in each molecule [congeries of hereditary particles] may admit of being discovered by noting the relative number of cases of each grade of deviation from the mean greyness.[26] [In Galton's example gray is the hybrid formed when equal amounts of black and white are mixed.]

Pascal's triangle, successive binomial expansions, would reveal the number of hereditary particles:

> If there were 2 gemmules only, each of which might be either white or black, then in a large number of cases one-quarter would always be quite white, one-quarter quite black, and one half would be grey. If there were 3 molecules, we should have 4 grades of color . . . and so on according to successive lines of 'Pascal's Triangle.'[27]

[23] Francis Galton, "Relative Supplies from Town and Country Families to the Population of Future Generations," *Journal of the Statistical Society of London*, Mar. 1873, *26*:19–26; "Proposal to Apply for Anthropological Statistics from Schools," *J. Roy. Anthro. Inst.*, 1873–1874, *3*:308–311; "Marlborough School Statistics," *J. Roy. Anthro. Inst.*, 1874, *4*:126–130; "On the Height and Weight of Boys Aged 14 in Town and Country Public Schools," *J. Roy. Anthro. Inst.*, 1876, *5*:174–180. Galton assumed that if the health of the school boys had degenerated, their breeding capacity would be impaired. His demonstrations were not conclusive (see *LLL* II 124), but this did not stop him (and subsequent eugenicists) from assuming that the hypothesis was correct.

[24] *Hereditary Genius*, p. 426.

[25] For the pangenesis experiments see *LLL* II 156–178, as well as Francis Galton, "Experiments in Pangenesis by Breeding from Rabbits of a Pure Variety . . . ," *Proc. Roy. Soc.*, 1870/1871, *19*:404; Charles Darwin, "Pangenesis," *Nature*, Apr. 27, 1871, p. 502; and Galton's reply, *Nature*, May 4, 1871, pp. 5–6.

[26] Galton to Charles Darwin, Dec. 19, 1875; *LLL* II 189–190.

[27] *Ibid.* R. C. Olby, *Origins of Mendelism*

This content downloaded from 165.123.34.86 on Thu, 15 Apr 2021 20:07:04 UTC
All use subject to https://about.jstor.org/terms

The fact that Galton was interested in binomial expansions tells us something significant about his model of heredity: the number of factors that contribute to the characteristics of any population is, in Galton's formulation, relatively small. A binomial curve represents the action of a *finite* number of definable causes; the right side of a binomial equation tells us how many different combinations can be produced, and with what frequency, when two factors a and b are combined in x (the exponent) ways. A normal curve, on the other hand, represents the action of an *infinite* number of undefinable small causes impinging upon any given phenomenon; the law of frequency of error does not try to examine what those causes might be, it simply postulates that they exist. Thus, by opting for binomial expressions Galton was opting for heredity over environment, nature over nurture. Environmental factors are infinite in number; organic factors can be restricted to the number of hereditary units in the germ cell. Exponential curves, according to Galton, "may occur in games of chance, but they assuredly do not occur in vital and social phenomena."[28]

Galton was wrong, and he soon realized his error, but the error, for our purposes, is instructive.[29] It demonstrates Galton's overriding concern with the eugenic ideal—the notion that environment has little influence on the character of men. Galton hoped for a binomial model of the phenomena of heredity, because a binomial model would have restricted the number of genetic determinants, and this would have implied that environmental factors are negligible. Eventually he was forced to discard the model, but he never discarded the eugenic conviction, as we shall see.

These papers in the early 1870s were Galton's first halting steps in the direction of biostatistics. The next step was not halting at all: it was an enormous leap and it occurred sometime between the publication of "Statistics by Intercomparison" in January 1875 and "Typical Laws of Heredity" in February 1877. Sometime during those twenty-four months Galton realized, as he said in his autobiography, that "the laws of heredity are solely concerned with deviations expressed in statistical units." The most likely date is the spring of 1876.[30] At that time Galton was puzzling over two sets of statistical data: one set from his sweet-pea experiments and the other from his inquiries into the height and weight of school boys.

The first experimental crop of sweet peas had been planted at Kew, in the spring of 1874, but that crop had failed. In order to avoid a second failure Galton sent sets of seeds (seven bags, containing ten seeds each, each group of ten containing seeds of the

---

(London:Constable, 1966), has suggested, on the basis of this letter, that Galton "had worked out the Mendelian explanation of hybridisation" (p. 76). This claim seems to me to be inflated. Galton's 1:2:1 ratio is phenotypical, referring to the physical appearance of the organism. Mendel realized that the appearance of the organism did not necessarily reflect its genetic constitution; his 1:2:1 ratio is genotypic—and the difference is crucial.

[28] Francis Galton, "Statistics by Intercomparison and Remarks on the Laws of Frequency fo Error," *Phil. Mag.*, Jan. 1875, *49*:34.

[29] Galton was wrong because normal and binomial curves are difficult to distinguish when the sample is high. Binomials are not useful for studies of large populations; there is no way to determine the probable deviation from a restricted number of observations. Many population characteristics obey normal distributions, and it is difficult to convert these to binomials, See Francis Galton, "Typical Laws of Heredity," *Proceedings of the Royal Institution*, Feb. 9, 1877. *8*:282–301.

[30] Pearson assumed that when Galton spoke of "the laws of heredity . . . solely concerned with deviations expressed in statistical units" he was referring to the discovery of correlation. Consequently Pearson placed this incident in 1888 (*LLL* II 393). The spring of 1876 seems more reasonable. As early as 1875 Galton had been using the expression "deviations expressed in

This content downloaded from
165.123.34.86 on Thu, 15 Apr 2021 20:07:04 UTC
All use subject to https://about.jstor.org/terms

same weight) to friends throughout the country and asked that the produce be sent to him in the same marked bags that had contained the parental seeds. Seven such crops succeeded, so that by the winter of 1875–1876 Galton had data from the produce of 490 seeds.[31] Similarly, by the same winter Galton had accumulated more data about the physique of boys in town and country schools: Eton, Haileybury, Wellington, City of London School, King Edward's School Birmingham, and Liverpool College had all sent returns.

What had Galton intended to do with all this data? For the school statistics that question is not difficult to answer: he wished to discover whether or not the physical character of school boys was deteriorating in the towns. But what of the sweet-pea data? Why had Galton gone to such lengths to determine the precise character and variability of sweet-pea plants produced by seeds of different sizes? The answer to that second question is related to the answer to the first. Galton took up the sweet-pea work because he realized that the school statistics could not yield the most important kind of eugenic information, information about the change in populations over several generations. Statisticians had observed that the bell curve for any character (let us say height) remains the same from one generation to the next; the filial and parental generations have the same mean value and the same variability spread.[32] If this is truly the case, what does it mean to say that populations are deteriorating or improving? Do abnormal parents produce normal offspring, despite their abnormality? This question could not be answered with the school statistics that Galton possessed, but a rough approximation of the answer could be garnered from plants grown under controlled conditions. "It was anthropological evidence that I desired, caring only for the seeds as means of throwing light on heredity in Man."[33] Several years earlier when Darwin suggested that Galton undertake plant breeding experiments in the interests of the theory of pangenesis Galton declined on the grounds that he had no facilities for gardening; when eugenic ideas were at stake he was capable of finding facilities with ease.[34]

During the winter of 1875–1876 Galton set about analyzing both the sweet-pea data and the school statistics. By the end of the spring of 1876 he realized that the law of deviations, which he had been using to derive information from the school statistics, could also be applied to the data from the plants. This was Galton's great imaginative

---

statistical units" to refer to the possibility of learning something about the character of a population without actually measuring it. (See "Marlborough School Statistics," *passim.*) By saying that "the laws of heredity are solely concerned with deviations expressed in statistical units," Galton simply meant that all of the hereditary processes had to be conformable to the law of frequency of error, just as the distribution of qualities throughout a population had to be similarly conformable. Galton said precisely this in "Typical Laws of Heredity": "The processes of heredity must work harmoniously with the law of deviation, and be themselves in some sense conformable to it" (p. 289). By "statistical units" Galton seems to have meant *any* statistical value—perhaps the mean and quartile. Pearson assumed that Galton was referring only to the probable error.

[31] Galton explained this procedure in "Typical Laws of Heredity," p. 290. For the failure of the first crop, see Galton to Charles Darwin, Apr. 14, 1875; *LLL* II 180.

[32] See, e.g., *Hereditary Genius*, Sec. 1, Ch. 3, where this phenomenon is discussed in detail.

[33] Francis Galton, "Address to the Anthropological Section of the British Association," *Brit. Assoc. Rep.*, 1885, p. 1207; also in *Nature*, Sept. 24, 1885, *32*:507. Galton was referring to his sweet-pea experiments when he made this remark.

[34] Galton to Charles Darwin, Mar. 31, 1870; *LLL* II 158–159.

This content downloaded from
165.123.34.86 on Thu, 15 Apr 2021 20:07:04 UTC
All use subject to https://about.jstor.org/terms

leap: the sweet-pea data could shed light on the anthropological data because they could both be made conformable to the same mathematical laws; heredity could be analyzed probabilistically.

The school statistics were presented to the public in May of 1876,[35] but the paper about the sweet peas—the paper which was to contain Galton's great insight into heredity—had to wait until the winter of the following year. The first version of this paper, which was entitled "Experiments with Plants on the Causes of Statistical Uniformity in Successive Generations," had been prepared by May 1876, but it was "burked" by George Howard Darwin:

> My dear George, How can I thank you sufficiently for the great trouble you have taken about my paper and your criticism. I will wholly, or almost wholly, rewrite and expand and not think of sending it in its present form.
> These confounded law of error ideas, which in themselves are so simple and clear but to express which no proper language exists and which lies so completely out of the every-day lines of thought are very baffling to deal with and to present, but I don't despair yet.[36]

"Typical Laws of Heredity" finally appeared in January of 1877; it was Galton's great breakthrough in the study of heredity. The data from the sweet-pea experiments revealed two peculiar phenomena: (1) the offspring of extreme plants are closer to the population mean than their parents are; (2) one degree of variation, one statistical unit, is the same absolute measurement whether the offspring of extreme parents or the offspring of normal parents are considered. These phenomena seem paradoxical when they are examined superficially—"although each individual does *not* as a rule leave his like behind him, yet successive generations resemble each other with great exactitude in all their general features"[37]—but Galton maintained that they become comprehensible when they are examined probabilistically, when they are examined in terms of the law of error.

Normal curves come in a variety of different shapes: they can be long and thin (indicating a small probable error and much clustering around the mean) or they can be rather squat (indicating a great range of variability in the population). If we assume that each step in the hereditary process is represented by a different curve, a population can revert to the mean and still produce offspring with the same range of variability as itself. Imagine that each individual is a pellet; if the mass of pellets together is shaped into a bell curve, then each grade of pellets can be dropped through a channel which, directed inward, will result in a bell curve which is longer and thinner than the original. Remove the bottom from the new bell curve–pellet mass, let the pieces fall freely, and the end result will be a curve of the same dimensions as the first. (See Fig. 1.) Thus a population may revert (fall toward the mean) and then vary (fall freely), but the end result will be a second population which is an exact reproduction of the first. Individual pellets may not fall into their original position (individuals may not be as tall as their parents), but the distribution of pellets will not have changed (there will be just as many giants as there were before). The phenomena of heredity can be explained without violating the properties of frequency curves.[38]

[35] Galton, "On the Height and Weight of Boys Aged 14," *loc. cit.*

[36] Galton to George Howard Darwin, May 17, 1876; Galton Archive, packet 6. No extant copy of the original paper has been found.

[37] "Typical Laws of Heredity," p. 283 (the italics are Galton's).

[38] *Ibid., passim.*

This content downloaded from
165.123.34.86 on Thu, 15 Apr 2021 20:07:04 UTC
All use subject to https://about.jstor.org/terms

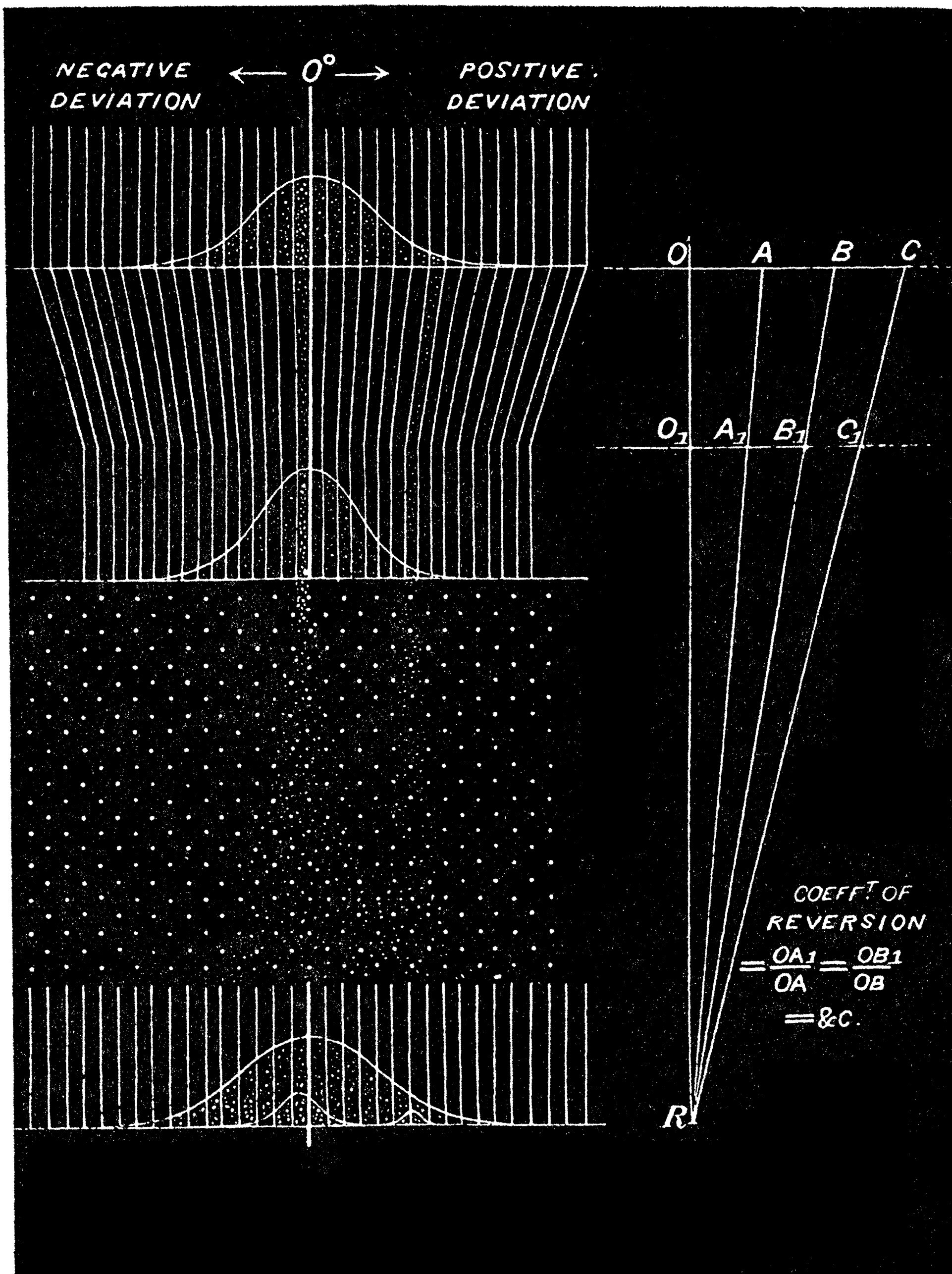

Figure 1. Galton's illustration of the nature of reversion, from "Typical Laws of Heredity," **Proceedings of the Royal Institution**, *Feb. 9, 1877.*

This content downloaded from
165.123.34.86 on Thu, 15 Apr 2021 20:07:04 UTC
All use subject to https://about.jstor.org/terms

Armed with a probabilistic model for heredity Galton realized that he needed more data of a very specific kind. If he knew, for example, the height of one thousand couples and all their offspring, he could calculate the amount of reversion that had occurred (by graphing each grade of parents against the mean height of their offspring), the extent of sexual selection (by the height of the father plotted against the height of the mother), whether variability had remained constant (is the probable error the same for the offspring of each grade?), and if it had not remained constant, how strong an influence had been wrought by natural selection and differential fertility. The riddle of hereditary stature would henceforth be solved, and he could go on to examine another characteristic, say weight or intelligence. If proper tests were constructed, the hereditability of every organic character would ultimately be known.

This is where the sweet peas came in: anthropometric data were hard to come by, but the sweet-pea data provided a good and readily available substitute. If the diameter of the mother seeds is plotted against the diameter of their offspring, can reversion be calculated? Galton drew a graph with parental seed size on one axis and filial size on the other; he marked the body of the graph wherever parental size and mean filial size intersected. When the marks were connected they formed a straight line with a measurable slope; in modern terminology, the first regression line had been drawn and the first regression coefficient had been measured.[39]

For Galton the slope of that line measured a fact of heredity: because every progenitor contributes to the heritage of the offspring, the offspring tend toward the mean of the race. For obvious reasons he called the slope the "coefficient of reversion." Within a decade he realized that this coefficient was not a product of heredity but of his own statistical manipulations; since "reversion" implied inheritance, Galton changed the name of the coefficient to "regression."[40] In 1888 he went one step farther: he realized that the same coefficient could determine the relationship between any two variates at all, whether or not they happen to be concerned with the hereditary process, whether or not they happen to be measured by the same absolute unit. At this juncture the "correlation" coefficient was born.[41]

The intellectual road that Galton traveled between the reversion coefficient of 1877 and the correlation coefficient of 1888 was far less convoluted than the one that had

[39] A copy of this diagram can be found in *LLL* IIIa 4; it comes from a notebook, entitled *Royal Institution Lecture*, which is in the Galton Archive. The text of the lecture ("Typical Laws of Heredity") as reprinted in the *Proceedings of the Royal Institution* does not indicate that the diagram was used.

Pearson believed that Galton first attempted to solve the reversion problem by graphing the *ranks* of parental size against the *ranks* of offspring size, subsequently drawing isograms to connect offspring of the same rank (*LLL* II 392–393). Pearson's suggestion on this matter was based upon his discovery of a demonstration diagram in which Galton attempted to illustrate precisely that procedure. Because Galton had used the words "average" and "quartile" in this diagram Pearson assumed that it must have been constructed during the 1870s, since Galton afterward changed those phrases to "mean" and "percentile." I doubt whether Pearson's presumption is correct; Galton continued to use "average" and "mean" interchangeably until at least 1889 (see Francis Galton, *Natural Inheritance*, London, 1889). "Quartile" was also used rather frequently until 1889 (*ibid.*, Ch. 5). Nonetheless, if Pearson is correct, his suggestion lends added credence to the idea that Galton's statistical work was heavily influenced by his experiences in meteorology; the isograms which Galton drew on that chart are identical to isobaric lines, and Galton was quite familiar with the manipulation of such lines.

[40] "Regression towards Mediocrity in Hereditary Stature," *J. Roy. Anthro. Inst.*, 1885, *15*: 246–263.

[41] "Co-relations and their Measurement, chiefly from Anthropometric Data," *Proc. Roy. Soc.*, Dec. 1888, *45*:135–145.

This content downloaded from
165.123.34.86 on Thu, 15 Apr 2021 20:07:04 UTC
All use subject to https://about.jstor.org/terms

taken him from the brute statistics of "Hereditary Talent and Character" to the sophisticated probability concepts of "Typical Laws of Heredity." After completing his analysis of the sweet-pea data Galton realized that the data were too sparse to be conclusive. He wanted better data; more specifically he wanted data from humans, not plants. "It was anthropological evidence that I desired, caring only for the seeds as a means of throwing light on heredity in Man."

From the schools Galton had acquired considerable information about the physique of children, but in order to measure the rate of reversion, he needed statistics from two *different* generations, and these he did not have. They continued to elude him until 1882 when he hit upon a brilliant solution to the problem: if anthropometric laboratories—offices for the weighing and measuring of human beings—were established in public places and parents were encouraged to visit these laboratories *with their children*, the records could be kept at a permanent institute and parents could be encouraged to send their children periodically for re-measurement.[42] Better yet, if each family kept its own personal record of its abilities and faculties, national statistics might be derived, as well as the medical history of the family itself.[43] Prizes might be given to physicians who kept the most capacious records of their patients. Prizes might even be offered to the families that kept the most complete records of themselves.[44]

> The advance of the science of heredity is seriously delayed through the want of such data. We do not know whether any given group of different characteristics which may converge by inheritance upon the same family will blend, neutralize or intensify one another, nor whether they will be metamorphosed and issue in some new form.... Our present ignorance of the conditions by which the level of humanity may be raised is so gross, that I believe if we had some dictator of the Spartan type, who exercised absolute power over marriages, assigning A to be the wife of B and C to be the wife of D, and who acted with the best intentions, he might possibly do even more harm than good to the race.[45]

Early in 1884 Macmillan's brought out the *Record of Family Faculties*, which summarized the information that Galton wanted; by May of that year 150 records had been filled and returned to him in London. During 1884 Galton also built and staffed an Anthropometric Laboratory at the International Health Exhibition at the South Kensington Science Museum. This laboratory, 6 feet wide and 36 feet long, contained instruments, as the introductory pamphlet described them, "for the Determination of Height, Weight, Span, Breathing Power, Strength of Pull and Squeeze, Quickness of Blow, Hearing, Seeing, Colour Sense and other Personal Data."[46] Here 9,337 persons were measured, each of them paying 3 shillings for the privilege. The data that Galton collected from this lab (it remained open in South Kensington for five years after the Health Exhibition itself had closed) became the raw material for his most important statistical discoveries.

[42] "The Anthropometric Laboratory," *Fortnightly Review*, 1882, N.S. *31*:332–338.

[43] "Medical Family Registers," *Fortnight. Rev.*, 1883, N.S. *34*:244–250.

[44] Galton placed an advertisement in the *Fortnightly Review* for Dec. 1883, offering prizes of up to £500 for the most complete family records; the names of the winners were printed in *Natural Inheritance*, pp. 75–77.

[45] Francis Galton, ed., *Record of Family Faculties* (London, 1884), p. 36.

[46] These instruments had been devised by Galton in cooperation with G. Croom Robertson. See *LLL* II 212 and 370; also Francis Galton, "Outfit for an Anthropological Laboratory" (London, privately printed, 1883). Many of the instruments that were used at South Kensington in 1884 are now in the Science Museum, Oxford University. A photograph of that lab can be found in *LLL* II 371.

This content downloaded from
165.123.34.86 on Thu, 15 Apr 2021 20:07:04 UTC
All use subject to https://about.jstor.org/terms

When the Health Exhibition closed in 1885 Galton came away with, among other things, data for the height of 928 offspring and their parents, just the sort of data that he needed to compute a reversion coefficient. Plotting the data was a tedious procedure, but it was not very complicated; the slope was 2/3. Galton was not surprised; the scantier sweet-pea data had earlier produced the same slope. Nor was he surprised to discover that when the process was reversed, when the values for individual offspring were plotted against the mean height of their parents, the coefficient was 1/3 and not the reciprocal of 2/3 as might have been expected. "The number of individuals who are nearly mediocre is so preponderant, that an exceptional man is more frequently found to be the exceptional son of mediocre parents than the average son of very exceptional parents."[47] In other words, it was not surprising to find that parents are 1/3 less deviate than their children, because most of the very deviate children would have become deviate not through normal heredity but by "sporting."

But Galton was surprised by something else. Rather meticulously he had drawn a chart in which the stature of every single offspring was entered opposite that of its mid-parent.[48] If, let us say, the mid-parental height had been 68 inches and if the height of the four children had been 64, 65, 66, and 67 inches, then a mark was placed in each of four vertical columns corresponding to those offspring heights on the horizontal line, 68 inches. Galton referred to this, and similar charts, as "Tables of Correlation."[49] In an effort to simplify the data Galton decided to count the number of entries in each square inch of the table. There were, for example, twelve offspring whose height was approximately 68 inches and whose mid-parental height had been approximately 69 inches. The results were entered at the intersections of the appropriate horizontal and vertical lines (see Fig. 2).

> I found it hard at first to catch the full significance of the entries, though I soon discovered curious and apparently very interesting relations between them. . . . Lines drawn through entries of the same value formed a series of concentric and similar ellipses. The common centre lay at the intersection of those vertical and horizontal lines which correspond to the value of $68\frac{1}{4}$ inches [the population mean for both offspring and parents]. . . . Their axes were similarly inclined. The points where each successive ellipse was touched by a horizontal tangent, lay in a straight line that was inclined to the vertical in the ratio of 2/3 and those where the ellipses were touched by a vertical tangent, lay in a straight line inclined to the horizontal in the ratio of 1/3.[50]

The elliptical diagrams were remarkable in themselves, but what happened next was even more striking to Galton. He realized that the diagrams could be constructed given only three pieces of information: (1) the probable error for the parental generation, (2) the probable error for the filial generation, and (3) the average reversion of the latter on the former. Using the mathematics of probability Galton thought it should be possible to redraw the same diagram without any additional data. He could not do this

[47] *Natural Inheritance*, p. 99. In *Natural Inheritance* Galton summarized all the statistical work that he had done between 1877 and 1888.

[48] A "mid-parent" is the average height of the two parents, after the mother's height has been transmuted into units that are probabilistically similar to the father's. See *Natural Inheritance*, pp. 5, 42, 87.

[49] *Memories*, p. 302.

[50] *Natural Inheritance*, pp. 101–102. In his autobiography Galton recounted the exact circumstances in which this discovery was made: "One morning, while waiting at a roadside station near Ramsgate for a train, and poring over the diagram in my notebook, it struck me, that the lines of frequency occur in concentric ellipses. The cases were too few for certainty, but my eye, being accustomed to such things, satisfied me that I was approaching the solution" (*Memories*, p. 302).

This content downloaded from
165.123.34.86 on Thu, 15 Apr 2021 20:07:04 UTC
All use subject to https://about.jstor.org/terms

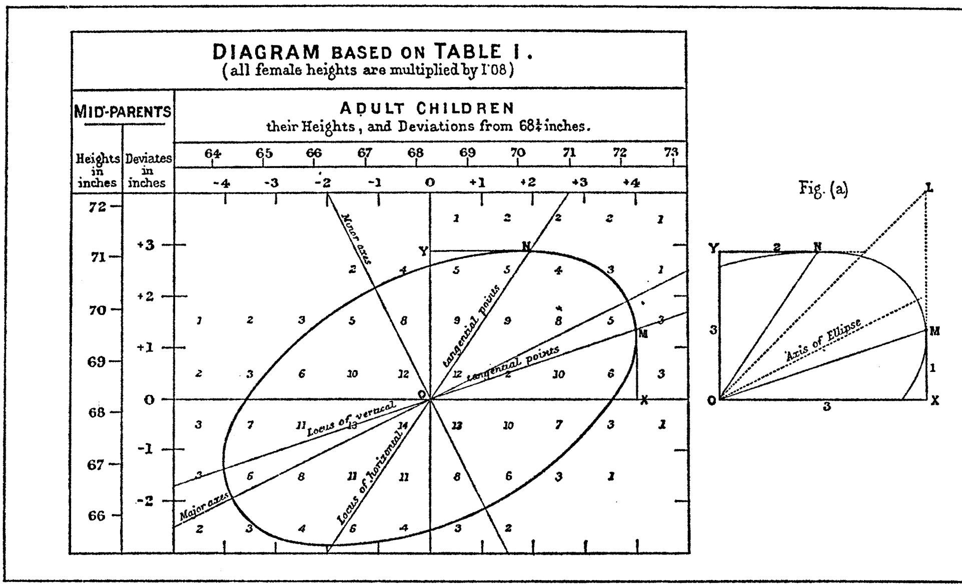

Figure 2. Galton's ellipse contour, drawn from his observations, as reproduced in Karl Pearson, **The Life, Letters and Labours of Francis Galton** *(Cambridge: Cambridge Univ. Press, 1914–1930), Vol. III, p. 14.*

himself because he had long since forgotten all the formulas for the conic sections, so he sent the problem to J. D. Hamilton Dickson, who was then Tutor of St. Peter's College Cambridge:[51] "I wrote down these [three] values, and phrasing the problem in abstract terms, disentangled from all reference to heredity .... I asked him kindly to investigate for me the Surface of Frequency of Error that would result from these three data, and the various shapes and other particulars of its sections that were made by horizontal planes."[52] Dickson's diagram was almost identical to the one that Galton had constructed.

> I may be permitted to say that I never felt such a glow of loyalty and respect towards the sovereignty and magnificent sway of mathematical analysis as when his answer reached me confirming, by purely mathematical reasoning, my various and laborious statistical conclusions with far more minuteness than I had dared to hope, for the original data ran somewhat roughly, and I had to smooth them with tender caution.[53]

Dickson had confirmed, deductively, the conclusions that Galton had reached empirically. Galton was delighted and surprised, surprised because of what Dickson's feat suggested: reversion occurs whenever two populations are compared in a probabilistic manner, *whether or not those populations happen to be connected genetically.* Reversion might well be a product of heredity, but it also seemed to be a product of statistical manipulation. For Galton this served as conclusive proof that the processes

[51] *Memories*, p. 302.

[52] *Natural Inheritance*, p. 102.

[53] "Address to the Anthropological Section of the British Association," *Nature*, Sept. 24, 1885, *32*:509.

This content downloaded from
165.123.34.86 on Thu, 15 Apr 2021 20:07:04 UTC
All use subject to https://about.jstor.org/terms

of heredity are probabilistic; it is no wonder that he was elated at Dickson's results. The name of the coefficient had to be changed; "reversion" was limited to heredity, "regression" would be biologically neutral.

His enthusiasm sparked, Galton began to play with the concept of regression. For many years he had been interested in measuring "degrees of family likeness;"[54] perhaps regression was the tool that he needed. Could he derive a different regression coefficient for each level of kinship? From the Anthropometric Laboratory data Galton calculated a regression coefficient for brother on brother: 2/3. Then he deduced a coefficient for uncles on nephews: "A nephew is the son of a Brother, therefore in this case we have $w = 1/3 \times 2/3 = 2/9$ [where w is the regression coefficient that is sought, 1/3 is the regression of sons on parents and 2/3 is the regression of brothers on each other]."[55] The regression of cousins on cousins was 2/27 and grandsons on grandparents, 1/9.[56] For "kinsman of any degree," once the distribution of height in the general population and the regression coefficient for any small fraternity are known, is $p^2w^2 + f^2 = p^2$, where p is the probable error for the whole population, f is the probable error for the distribution of ranks of height in the population of kinsmen to be calculated, and w is the regression coefficient for any other part of the population.

The prospects were too enticing; Galton could not resist going further. Could he calculate the contribution of each ancestor to the heritage of any single individual? Proceeding from a set of false assumptions through a sequence of fallacious mathematical processes, Galton arrived at his law of ancestral inheritance:[57] each parent contributes 1/4 of the heritage, each grandparent 1/8, each great-grandparent 1/16, and so on.[58]

Galton was still not content to stop. Would the regression law work for characters which do not "blend" in quite the way that stature does? What does it mean to say, for example, that offspring are 2/3 less deviate with respect to eye color? Is there an average value between blue and black eyes? Is there a population mean for eye color? Can offspring regress toward it? At the very least Galton thought that it might be possible to find out whether distribution of eye color had been changing (whether, in other words, there were more dark-eyed people in one generation than there were in another) and whether the eye color of children could be predicted from that of their parents, using the law of ancestral inheritance. The mathematics is somewhat dubious, but Galton did conclude that the distribution of eye color, like that of stature, had remained constant from one generation to the next; this implied that the probable eye color of children of known parents could be predicted.[59] Galton also believed that if

[54] Composite portraiture was one of the techniques he had developed for measuring family resemblance. See *LLL* II Ch. 12; also Francis Galton, "Composite Portraits made by combining those of many different Persons into a single resultant Figure," *J. Roy. Anthro. Inst.*, 1878, *8*: 132–142.

[55] *Natural Inheritance*, p. 132. The technique was first expounded in "Family Likeness in Stature," *Proc. Roy. Soc.*, 1886, *40*:42–73.

[56] See *Natural Inheritance*, p. 133, for a chart detailing all these familial relationships.

[57] *LLL* IIIa 22–23; R. G. Swinburne, "Galton's Law-Formulation and Development," *Annals of Science*, 1965, *21*:15–31.

[58] Galton had actually assumed the existence of this sort of law as far back as 1865 ("Hereditary Talent and Character," p. 322), and Darwin had made the same assumption in *The Variations of Animals and Plants* (Vol. II, pp. 34, 88). One has the faint suspicion that Galton was trying to demonstrate a principle that he believed intuitively.

[59] "Family Likeness in Eye Colour," *Proc. Roy Soc.*, 1886, *40*:402–416; also *Natural Inheritance*, Ch. 8.

This content downloaded from 165.123.34.86 on Thu, 15 Apr 2021 20:07:04 UTC
All use subject to https://about.jstor.org/terms

the regression technique could be used on eye color, it also could be used to analyze the distribution and inheritance of such psychological qualities as artistic ability and legislative skill.[60]

Some of these exercises may have been fallacious, some naïve, and some absurd, but no reader could have come away from *Natural Inheritance* (1889), the book in which they were all summarized, without being struck by the manifold ways in which the frequency-of-error laws could be, and had been, manipulated. Some of the results may have been dubious and some of them may have been more or less useless, but the possibilities of the technique which had been demonstrated—its versatility and investigative power—these must have been awe inspiring. Galton himself was overwhelmed: "The Law of Frequency of Error," he said, "would have been personified by the Greeks and deified, if they had known of it. It reigns with serenity and in complete self-effacement amidst the wildest confusion."[61] Karl Pearson, who was then a lecturer in engineering at University College, London, read *Natural Inheritance* soon after it was published and immediately became a convert; he delivered his first lecture on Galton's statistical method in March 1889.[62] The book, and the statistical technique which it had propounded, caused a philosophical crisis in Pearson's life:

> Galton wrote to Darwin on December 24, 1869 that the appearance of the *Origin of Species* had formed a real crisis in his life and freed him from his old superstition as if he had been roused from a nightmare.... For some of us Galton's new calculus acted in precisely the same manner; it enabled us to reach real knowledge ... in many branches of inquiry where opinion only had hitherto held sway. It relieved us from the old superstition that where causal relationships could not be traced, there exact or mathematical inquiry was impossible. We saw the field of scientific, or quantitative, study carried into organic phenomena and embracing all the things of the mind. It was for us the dawn of a new day.[63]

One more step remained for Galton; one more extension of the regression idea remained to be made. The problem of measuring family resemblances had been, as we have seen, part and parcel of his interest in heredity and eugenics. Even after 1889 he continued to search for better solutions to the problem, but they eluded him. In 1906 he wrote: "I am hard at work on my often-taken up subject 'Measurement of Resemblance'.... I have bungled it to an extraordinary extent, in tangled bye-paths before finding the direct way."[64] The subject had indeed been taken up far too often for comfort; the bunglings had been numerous and the by-paths had been very frustrating.

One of those tangled by-paths had been Bertillon's technique of personal identification, and it was this track which led Galton to the idea of correlation. Alphonse Bertillon, a French prison official, had developed a technique for identifying prisoners which consisted of measurements of the head and limbs. Having measured head length, head breadth, foot length, and middle-finger length of the left foot and hand, Bertillon thought that the combination of measurements would form a unique anthropometric profile of each individual. In 1888 Galton was asked to lecture on the subject at the Royal Institution. His initial reaction was positive; Bertillon's system might, he thought, reveal "independent features suitable to hereditary investigation ... making

[60] *Natural Inheritance*, Ch. 9.

[61] *Ibid.*, p. 66.

[62] See Karl Pearson, "W. F. R. Weldon," *Biometrika*, 1906, *4*:n., p. 16.

[63] *LLL* II 357–358.

[64] Galton to his sister, Elizabeth Galton Wheler, Jan. 30, 1906; Galton Archive, box 35.

This content downloaded from
165.123.34.86 on Thu, 15 Apr 2021 20:07:04 UTC
All use subject to https://about.jstor.org/terms

it possible to trace kinship with considerable accuracy."[65] But as he considered the matter in the months following his lecture at the Royal Institution he found serious objection. Bertillon assumed that all the measured qualities were independent variables, but Galton realized that in point of fact they were not: head breadth is dependent on head length, and foot length is dependent on the length of the middle toe. A unique personal profile could not be formed under such conditions; long heads would go together with narrow heads, and what one would have, more often than not, would be a description of a "type," not a person.[66]

Galton wondered just how fallacious Bertillon's system might be: could he determine the dependence of each measurement upon the others? The idea of organic correlation was hardly new. In *The Variations of Animals and Plants* Darwin had written, "All the parts of the organization are to a certain extent connected or correlated together," and Galton had underlined the words.[67] Now Galton proposed to investigate the extent of that connection. He arranged his data (Bertillon's statistics as well as limb measurements that had been taken in his own Anthropometric Laboratories) just as he had arranged the data for stature in fathers and sons: arm length, for example, was plotted on one side of a chart, and height, let us say, on another. If a man's height was 69 inches and his arm length 35 inches, an entry was made at the intersection of the two appropriate lines.

> Not only did it turn out that the measure of correlation between two variables is exceedingly simple and definite, but it became evident almost from the first that I had unconsciously explored the very same ground before. No sooner did I begin to tabulate the data than I saw that they ran in just the same form as those that referred to family likeness in stature. . . . A very little reflection made it clear that family likeness was nothing more than a particular case of the wide subject of correlation, and that the whole of the reasoning already bestowed upon the special case of family likeness was equally applicable to correlation in its most general aspect.[68]

Galton had discovered that regression coefficients occur even when the distributions being compared are not "populations" in any ordinary sense of the word. He also had discovered that the dependence of one factor upon another could be measured, *even if the two factors were not commensurable*. Lung capacity could be correlated with chest breadth despite the fact that the two are measured in different units; all he needed was the measurements expressed in standard units, in degrees of deviation. If an individual had a chest breadth one unit deviant from the norm (one unit of probable error) and a lung capacity which was two units deviant, a mark could be made for him at the intersection of the two appropriate lines on a chart. Ellipses could be drawn through the intersections with equal numbers of cases (just as they had been drawn for the stature data), tangents to these ellipses could be found, the slopes of the tangents could

[65] Francis Galton, "Personal Identification and Description," *Nature*, June 21, 1888, *38*:202.

[66] Francis Galton, "Presidential Address to the Anthropological Institute," *J. Roy. Anthro. Inst.*, 1888, *19*:401–419. Galton's interest in fingerprints arose out of his dissatisfaction with Bertillon's system of personal identification. See Francis Galton, "Identification by Finger-Tips," *Nineteenth Century*, Aug. 1901, *30*:303–311; also "The Patterns in Thumb and Finger Marks," *Philosophical Transactions of the Royal Society*, 1891, *186*:1–23; and *Finger Prints* (London, 1893). By the turn of the century fingerprinting hnd become the common technique for identifying criminals in England.

[67] *Variations*, Vol. II, p. 319. Galton's copy of the book is in the Galton Library, Statistics Department, University College, London.

[68] "Presidential Address to the Anthropological Institute," p. 411.

This content downloaded from
165.123.34.86 on Thu, 15 Apr 2021 20:07:04 UTC
All use subject to https://about.jstor.org/terms

be measured, and finally one would have a regression—no, a correlation—coefficient, a measurement of the extent to which lung capacity is dependent upon chest breadth.[69]

Galton summarized the correlational calculus in a short paper which he sent to the Royal Society in December 1888.[70] From Pearson's point of view that short paper was the most important of all Galton's writings:

> Thousands of correlation coefficients are now calculated annually; the memoirs and text-books on psychology abound in them; they form the basis of investigations in medical statistics, in sociology and anthropology.... Galton's very modest paper of ten pages has led to a revolution in our scientific ideas. . . . Formerly the quantitative scientist could only think in terms of causation, now he can think also in terms of correlation. This has not only enormously widened the field to which quantitative and therefore mathematical methods can be applied, but it has at the same time modified our philosophy of science and even of life itself.[71]

Pearson wrote in 1930; the intervening decades have given us no reason to doubt his estimate of Galton's work.

Oddly enough Galton never explored the possibilities of the technique he had discovered. In his autobiography he devoted only one page to correlation and did not indicate that he thought of it as a major contribution to science.[72] After 1888 he used the technique on only one or two occasions, in particular in his attempts to discover whether fingerprint patterns are correlated in members of the same family or on the several digits of the same individual. By 1888 Galton had probably reached the limits of his investigative vigor; he was then sixty-six years of age. The remaining two decades of his life he devoted to popularizing the fingerprint technique, propagandizing for eugenics, and generally enjoying the modest scientific fame he had acquired. During those years he encouraged, stimulated, and often financed the work of a younger group of "passionate statisticians" who were fascinated by his method: Pearson, W. F. R. Weldon, Edwin Schuster, and others. Perhaps he felt that he should leave it to them to develop the possibilities of correlation. Perhaps he simply did not realize the enormity of what he had done—it is hard to tell.

Regression and correlation were Galton's two most important contributions to the history of the sciences. These statistical ideas were founded on the rock of Galton's interest in heredity. Their very names give us clues to their origin: "regression" was once "reversion," and reversion was interesting to Galton because he believed it was part of heredity. "Correlation" has two roots: first as "co-relation," a measure of the coincidence of two variables, and second as "correlation," the biological principle which relates the growth of one part of the body to the growth of another. Galton believed, as Darwin had believed, that correlation is also an aspect of heredity.[73]

Galton created biostatistics while he was in pursuit of a solution to the problem of heredity. He dreamed of a truly eugenic society, a society based upon the laws of heredity: the laws of heredity would guide the breeding habits of men, and the evolutionary welfare of the race would become a moral criterion. The title of Galton's

[69] See Francis Galton, "On the Principle and Methods of Assigning Marks for Bodily Efficiency," *Brit. Assoc. Rep.*, 1889, 474–478; also *Nature*, Oct. 31, 1889, *40*:650–651.

[70] "Co-Relations and their Measurement, chiefly from Anthropometric Data," *Proc. Roy. Soc.*, Dec. 1888, *45*:135–145.

[71] *LLL* IIIa 56–57.

[72] *Memories*, p. 302.

[73] See Darwin, *The Variations of Animals and Plants*, Vol. II, pp. 319–338.

This content downloaded from
165.123.34.86 on Thu, 15 Apr 2021 20:07:04 UTC
All use subject to https://about.jstor.org/terms

Herbert Spencer Lecture in 1906 is actually a summary of his life's work: "Probability —the Foundation of Eugenics."[74] Probability techniques were the only tools that could unlock the secrets of heredity; the secrets of heredity were the bases of eugenics. Galton might also have entitled his lecture "Eugenics—the Foundation of Biostatistics," because his eugenic dreams had provided him with the motivation and the mental perseverance that he needed to unlock the secrets of probability.

[74] Published at Oxford (Clarendon Press, 1907). See *LLL* IIIa 317–323.

This content downloaded from
165.123.34.86 on Thu, 15 Apr 2021 20:07:04 UTC
All use subject to https://about.jstor.org/terms