Studies in History and Philosophy of Science 86 (2021) 6–19

Contents lists available at ScienceDirect

Studies in History and Philosophy of Science

journal homepage: www.elsevier.com/locate/shpsa

Francis Galton's regression towards mediocrity and the stability of types

Adam Krashniak*, Ehud Lamm

The Cohn Institute for the History and Philosophy of Science and Ideas, Tel Aviv University, Ramat Aviv, 6997801, Tel Aviv, Israel

ARTICLE INFO

Article history:
Received 13 August 2020
Received in revised form
29 October 2020
Accepted 27 December 2020
Available online 16 January 2021

Keywords:
Francis Galton
Regression to the mean
Inheritance and statistics
Stability of types
Evolutionary saltationism

ABSTRACT

A prevalent narrative locates the discovery of the statistical phenomenon of regression to the mean in the work of Francis Galton. It is claimed that after 1885, Galton came to explain the fact that offspring deviated less from the mean value of the population than their parents did as a population-level statistical phenomenon and not as the result of the processes of inheritance. Arguing against this claim, we show that Galton did not explain regression towards mediocrity statistically, and did not give up on his ideas regarding an inheritance process that caused offspring to revert to the mean. While the common narrative focuses almost exclusively on Galton's statistics, our arguments emphasize the anthropological and biological questions that Galton addressed. Galton used regression towards mediocrity to support the claim that some biological types were more stable than others and hence were resistant to evolutionary change. This view had implications concerning both natural selection and eugenics. The statistical explanation attributed to Galton appeared later, during the biometrician-mutationist debate in the early 1900s. It was in the context of this debate and specifically by the biometricians, that the development of the statistical explanation was originally attributed to Galton.

1. Introduction

Francis Galton (1822–1911) was a scientific polymath who made pioneering contributions to several fields, including the study of heredity in populations and the statistical analysis of data. Motivated by eugenic concerns, he studied patterns of human heredity. It is often claimed that while undertaking this work he discovered the statistical phenomenon of regression to the mean and that this discovery was a pivotal moment in the history of statistical thinking. In this paper, we argue that paying close attention to Francis Galton's views about heredity and to his anthropological concerns compels us to revisit his role in the history of the concept of regression to the mean as it is often described in the history of biology and statistics. The historical account we reject foregrounds the development of statistical thinking over Galton's other pursuits.

Francis Galton was born to a Quaker family of the upper social classes in Britain and was designated to become a physician like his grandfather Erasmus Darwin (Porter, 1986, p. 131). At age 18, with the encouragement of his older cousin Charles Darwin, he quit his medical studies and went on to study mathematics at Cambridge (Gillham, 2001, p. 31). During his time in Cambridge he suffered from two mental breakdowns preventing him from finishing his degree (Porter, 1986, p. 131). After leaving Cambridge, he decided to travel, the considerable inheritance he received after his father's death freeing him from the need to make a living. In the years that followed he turned his journeys into scientific endeavors. Galton joined expeditions around Africa, and made numerous contributions to the Royal Geographical Society (Porter, 1986, p. 132; Gillham, 2001, ch. 5–10). These expeditions seem to have made a lasting impact on his views on racial differences and human psychology (Fancher, 1983; Stocking, 1987). He also conducted meteorological research, which may have inspired his later statistical work (Cowan, 1972, pp. 513–4; Porter, 1986, pp. 272–3). From around 1865 onwards, Galton began studying the processes of heredity as well as variations in human traits. As we shall see in the next section, the underlying theme that connected most of these various works were his eugenic motivations: his deep conviction that differences between people, mental as well as physical, were hereditary, and his attempts to understand the processes by which traits were inherited. These works led to Galton's development of the concepts of regression and correlation, as well as to his conclusions regarding inheritance and evolution.

This paper will criticize what has become a common narrative regarding Galton's work on regression towards mediocrity.

* Corresponding author. The Cohn Institute for the History and Philosophy of Science and Ideas, Tel Aviv University, Ramat Aviv, 6997801, Tel Aviv, Israel.
E-mail address: adamkrashniak@gmail.com (A. Krashniak).

https://doi.org/10.1016/j.shpsa.2020.12.008
0039-3681/

Regression towards mediocrity was the fact that offspring of parents who deviated significantly from the mean tended on the average to deviate less than their parents. On the narrative we argue against, Galton's explanation for this phenomenon changed through his career from a biological explanation to a statistical one (Hacking, 1990; Pearl & Mackenzie, 2018; Stigler, 1999, 2016; Witteveen, 2019). In his earlier works on heredity, Galton explained this phenomenon by appeal to an inheritance process that caused offspring of parents who deviated significantly from the mean trait to revert back to the mean and hence deviate less than their parents (e.g., Galton, 1877). However, after conducting his analysis on stature in human families in the years 1884–5, and by the time he wrote *Natural Inheritance*, published in 1889, Galton came up with a new way to explain regression towards mediocrity, as a statistical phenomenon, now known as "regression to the mean". The novelty of this statistical explanation was that it showed that the fact that offspring deviated from the mean less than their parents was a necessary consequence of other statistical properties of the parent and offspring populations, namely their distributions and the correlation between them, without referring to the specific causal details of inheritance (for a discussion of statistical explanations, see Ariew et al., 2015). This explanation also showed that the fact that offspring deviated from the mean less than their parents did not imply the existence of an inheritance process that caused offspring to revert to the mean. As such, it led Galton to abandon his earlier ideas regarding the existence of such an inheritance process.

In contrast with this common historical narrative, we will argue that Galton never changed his explanation of regression towards mediocrity from a biological to a statistical one. We examine the evidence given to support the claim that Galton explained regression towards mediocrity statistically, and argue that the evidence fails to show this. Moreover, we show that Galton never abandoned his ideas concerning a process of inheritance that caused offspring to deviate less significantly from the mean than their parents, but rather continued to explain regression towards mediocrity by appealing to just such an inheritance process throughout his scientific career. Our analysis of Galton's work on regression towards mediocrity addresses Galton's statistical ideas as well as his ideas about inheritance and evolution, in contrast with the common narrative which focuses almost exclusively on Galton's statistics.

Galton's work on regression towards mediocrity served his larger biological and evolutionary concerns. Associating regression towards mediocrity with an inheritance process that caused offspring to deviate less significantly from the mean than their parents played an important part in Galton's ideas regarding the stability and immutability of biological and human groups. Regression towards mediocrity also played an important role in Galton's ideas that some biological traits could be referred to as the real type of a population; an idea that had clear implications for discussions that were common at the time about the differentiation and classification of different biological and human groups.

We will also show that understanding the statistical explanation of regression to the mean indeed served as an argument against the existence of the kind of inheritance process described by Galton; however, this development did not take place in Galton's work on statistics and inheritance at the end of the 19th century. Rather, the statistical explanation of regression to the mean was developed in the first years of the 1900's, as part of the debate between Karl Pearson and Raphael Weldon of the biometric school and William Bateson and Hugo De Vries of the mutationist-mendelian school on whether evolution proceeded in gradual steps or in big jumps. While Bateson and De Vries adopted Galton's association of regression with the existence of an inheritance process that caused offspring to revert back to the mean, Pearson and Weldon used the statistical explanation of regression to the mean to undermine this mutationist idea. Furthermore, the statistical explanation of regression to the mean was originally attributed to Galton by Pearson and Weldon, as part of this debate with the mutationists. Thus, on our telling, Galton's work on regression towards mediocrity shows more continuity between Galton's ideas on regression and those of the mutationists and less continuity between Galton's ideas on regression and those of the biometricians than is recognized. In addition, our interpretation shows that the common narrative misses the gradual development of the statistical concept of regression to the mean and the scientific context in which it was developed.

2. Galton's path to regression

2.1. Eugenic motivations and the law of deviation

Several historians have emphasized eugenics as the guiding motivation for Galton's development of statistical tools to study inheritance (e.g., Cowan, 1972; Mackenzie, 1981; Porter, 1986; Desrosières, 1998; but see Waller, 2004). Approaching middle age, and after making notable contributions to fields such as geography and meteorology, Galton read Darwin's *On the Origins of Species*, published in 1859. As Galton described in his autobiography, reading Darwin's book "made a marked epoch in my own mental development" and encouraged him to pursue "many inquiries which had long interested me, and which clustered round the central topics of Heredity and the possible improvement of the Human Race" (Galton, 1908, pp. 287–8). Galton was occupied for many years with the apparent differences between human beings in their talent and abilities, and he was inclined to believe that those differences were hereditary (Gillham, 2001, p. 155). Darwin's work led Galton to reject at last any commitment to Christian or "supernatural" ideas and turn to a naturalistic worldview which conceived of the mental traits of men as being determined by nature and inherited just like physical traits of men and animals (Mackenzie, 1981, pp. 54–5; Porter, 1986, pp. 132–3). Hence, these traits could be improved by man himself in a process similar to what Darwin termed "artificial selection", in which those with the best traits were chosen to breed.

According to Donald Mackenzie, Galton's ideas about eugenics reflected the interests of the social group that he belonged to, the British professional middle class (Mackenzie, 1981, pp. 52–6). By arguing that talent and ability were inherited, and that eminent men would succeed regardless of the social class they were born into, Galton "legitimated the elevated position of the professional elite to which he belonged", showing that it was a "natural elite, not merely a social one" (Mackenzie, 1981, p. 53). The naturalism conveyed by Galton's eugenic ideas was also a weapon of the middle class against the authority of the church and its monopoly on many professional positions (Mackenzie, 1981, pp. 54–5). By denying the supernatural, Galton argued for the higher authority of the scientific profession to which he and his likes belonged, and for the need for a 'scientific priesthood' which would replace the traditional one. Theodore Porter argued that Galton's arguments against the authority of religion and the church, which were based on his eugenics, reflected his attitude as a deeply committed social reformer who aimed to reconstruct society in the image of science (Porter, 1986, pp. 130–1, 133).

Galton's thesis that human talent and ability were hereditary appeared in his first article on heredity, which was published in 1865 in *Macmillan's Magazine* (Galton, 1865), as well as in his first book on the subject published four years later, *Hereditary Genius* (Galton, 1869). The practice of measuring human traits and studying their distributions was common at the time. Galton referred in *Hereditary Genius* to Adolphe Quetelet's influential analysis of chest measurements of 5,378 Scottish soldiers, published in 1844, which concluded that chest sizes had a bell shaped distribution (Galton, 1869, pp. 29–30). The eminent German scientist Rudolf Virchow (1821–1902) organized a large scale survey of schoolchildren in the 1870s for the German Anthropological Society (nearly 7 million children were surveyed), though the traits he studied were not continuous.[1] Similar studies were later conducted in Austria, Switzerland, Holland, and Belgium, though a lot of data were collected more haphazardly by, for example, seizing on travelling shows and fairs. Development of accurate measuring tools and standards was the order of the day. Much of this work was concerned with the possibility of classifying humans into types and determining the mutability of these types, typically considered through the ambiguous concept of race. While the French biologist Jean Louis de Quatrefages (1810–1892) thought that members of the same nation belonged to a single race, Virchow used his work on races in Germany to argue to the contrary and concluded that nations were composite in character. He thus favored talk of human stocks rather than races and also, unusually for the period, concluded that skull anatomy was plastic.

Galton's anthropometric work was thus part of a rich vein of work on human types and races that was concerned with questions of social evolutionism and progress. Paul Broca (1824–1880), the founder of the Anthropological Society of Paris, who influenced Virchow's work, collected brains of animals and eminent humans to compare their sizes. The Belgian anatomist Anders Retzius (1796–1860) developed a metric called the cephalic index that was used to compare skull shapes and to map their variation across Europe. Several decades later, in the context of ongoing eugenic debates concerning immigration laws and restrictions in the United States, Franz Boas1858–1942), who crossed paths with Virchow in the Berlin Ethnological Museum in 1883–1886, and knew Galton's methods (Boas, 1893), would apply the same general methodology to immigrants to the United States. He measured close to 18,000 individuals and concluded, echoing Virchow's non-hereditarianism to an extent, that the skull shapes of immigrant children resembled those of their adopted country rather than their parental stock. We can see that questions about the reality of types, about their stability and heritability, and about whether populations were mixtures and how this was to be determined, as well as debates about the conclusions to be drawn concerning the history of humans, were prevailing concerns. So were questions about the implications for public policy.[2] Galton was not unique in addressing these questions, but rather part of a broader discussion. In *Hereditary Genius* Galton appealed to a notion of stability of types drawn from mechanics to ground his discussion (Galton, 1869, p. 369); the very same metaphor is echoed in his later book *Natural Inheritance*. Unchanged by his alleged discovery of regression to the mean, Galton's arguments towards the end of his career concerning the stability of types would continue to address the prevailing concerns of the day; the same concerns that animated his earlier studies.

In both his 1865 essay and in *Hereditary Genius*, Galton aimed to establish the thesis that talent was hereditary by collecting data on eminence of men (e.g., by surveying biographical works on notable men (Galton, 1865)), and showing based on these data that sons of eminent men had greater chances of being eminent than sons of ordinary men. Galton sought a method for quantifying the differences in eminence between men, and in *Hereditary Genius* he offered such a method, based on the law of deviation.

The law of deviation (or law of error) described the mathematical properties of a bell shaped distribution, and explained how this distribution emerged as a result of a host of accidental causes (this explanation is described in more detail below).[3] The development and use of the law, originally in the physical sciences and later on in the social sciences, took place mainly on the Continent (Mackenzie, 1981, p. 57; Stigler, 1986). Galton became acquainted with the law around 1860, first when his old friend the geographer William Spottiswoode explained the law to him, and later when he encountered it in the work of Quetelet (Galton, 1908, p. 304; Gillham, 2001, pp. 157–8).

In *Hereditary Genius*, Galton argued that since many physical traits of men were shown to distribute in a bell shaped fashion, this had to be the case for talent and ability as well (Galton, 1869, p. 26; 31–2). This enabled him to talk about the proportion of men who had average, or mediocre, ability, as compared to the proportion of eminent men, and to quantify just how much the ability of eminent men differed from that of mediocre men (Galton, 1869, pp. 33–6; Hilts, 1973, p. 224). In the following years he tried to develop a statistical method for expressing the patterns of heredity of traits from parents to offspring based on the law of deviation (Hilts, 1973, p. 225; Stigler, 1986, pp. 272–5). These attempts culminated with the work he presented in 1877.

2.2. Reversion and regression

On the 9th of February, 1877, Galton gave a lecture at the Royal Institution in London, titled "Typical Laws of Heredity" (Galton, 1877). The aim of the lecture was to present a problem that arose from what seemed like two contradicting observations about the patterns of inheritance, and to offer a solution to the problem. Galton pointed to many records which showed that the distributions of traits in biological populations tended to remain constant over generations (Galton, 1877, p. 282). This would not have come as a surprise if parents gave birth to offspring that were identical to them, but that, Galton noted, was not often the case (Galton, 1877, pp. 282–3). Hence, Galton posed the problem as follows: "How is it, that 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?" (Galton, 1877, p. 283).

Galton noted that in addition to being constant over generations, the distributions of biological traits also conformed to the law of deviation and exhibited a bell shaped pattern (Galton, 1877, p. 283), a fact he had already noted in *Hereditary Genius* eight years earlier. This observation allowed Galton to approach the problem of the inter-generational stability of trait distributions by using a new device, later named the quincunx, which was built by him in order to explore the law of deviation.

The quincunx was probably devised by Galton in 1873 (Stigler, 1986, p. 277). It consisted of a vertical board in which rows of pins were fixed (Fig. 1). The apparatus had an opening at the top center, from which balls made of lead could be dropped. The balls collided with the pins as they fell down, causing them to end up in one of several compartments at the bottom of the device. Galton originally used the quincunx to demonstrate the law of deviation, by showing how a bell shaped distribution emerged as a result of many small accidental causes (or "a host of petty disturbing influences" as Galton referred to them (Galton, 1877, p. 289)). According to the prevailing interpretation of the law of deviation, the mean value was the result of a constant cause, while independent

[1] For the context of this work and its reception see (Ackerknecht, 1953, p. 211).

[2] See (Ackerknecht, 1953; Cryle & Stephens, 2017; King, 2019; Porter, 2018; Stocking, 1987).

[3] The bell shaped distribution will be referred to as the "normal" distribution only later by Galton himself in his book *Natural Inheritance* (e.g. Galton, 1889b, p. 54).

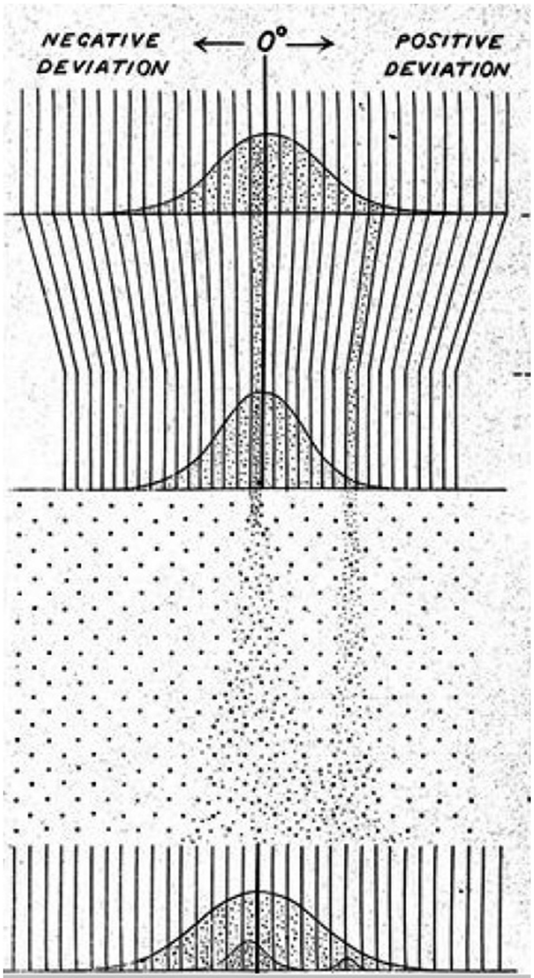

Fig. 1. Galton's modified quincunx (Galton, 1877). The bottom part of the figure represents the operation of Galton's original quincunx.

causes with equal intensity, acting indifferently in all directions, were considered as accidental causes that produced a bell shaped distribution around the mean (Galton, 1877, p. 289; Hilts, 1973, pp. 209–10, 216–7; Stigler, 1986, p. 274). The pins in the quincunx represented such accidental causes because whenever a ball hit a pin it had an equal chance of falling to the left or to the right of the pin, and these chances were independent of the effects that other pins had on the trajectory of the ball (Galton, 1877, p. 289). When many balls were dropped into the device, a bell shaped curve emerged, with most balls ending in the middle compartment, fewer balls ending in compartments to the left or to the right, and yet fewer balls ending in compartments that were further away from the middle one.

In 1877, Galton used the quincunx for a novel purpose. He argued that explaining the bell shaped distribution as the result of a host of petty influences could not explain how the bell shaped distribution of traits in biological populations remained constant over generations (Galton, 1877, pp. 289–90). Biological traits, Galton remarked, were passed on from one generation to the next by the processes of inheritance, which were not a set of accidental causes. He also argued that earlier thinkers who measured biological traits and showed that they conformed to the law of deviation, such as Quetelet, completely overlooked this problem (Galton, 1877, p. 289). His goal then, was to show how the processes of inheritance "work harmoniously with the law of deviation, and be themselves in some sense conformable to it" (Galton, 1877, p. 289). Galton set forth to show this by proposing a modification of the quincunx.

The modified quincunx consisted of two partitions (Fig. 1). In the top partition, balls fell through rows of pins to the bottom compartments of a quincunx. In the bottom partition, they were then released through small doors at the bottom of each compartment for another round through a quincunx (we do not know if this apparatus was ever built). Each ball represented a biological parent, the compartment where the ball was located represented the parent's trait value (for instance, balls in the middle compartment represented individuals with average height while balls in the left and right compartments represented taller and shorter individuals), the falling of the balls from their locations to a compartment below represented the process of inheritance, and the new location of the ball in the lower compartment represented the trait of an offspring. Galton intended this arrangement to represent inheritance in the following way: the first round through the quincunx produced a bell shaped distribution representing a parent population, while the next round through the quincunx was intended to represent the trait values in the population of their offspring.

Galton next introduced two processes of inheritance that purportedly worked together to keep the distribution of traits constant and bell shaped across generations (Galton, 1877, pp. 291–2). The "family variability" process of inheritance was demonstrated by opening the trap door under each of the upper level compartments in the quincunx and letting the balls fall through the rows of pins producing a bell shaped distribution centered around the compartment that was directly below it. This process increased the dispersion of the population, and had it been the only process of inheritance that took place, the distribution of traits in populations would become more dispersed with each generation.

The second process Galton introduced, which he called reversion, counteracted the increasing dispersion produced by family variability (Galton, 1877, p. 292). The balls in the upper level compartments in the quincunx now first fell down through funnels before reaching the compartment doors from which they were released to collide with the pins in the device. Balls that were located in the middle compartment fell straight down through the funnels while balls located to the right or to the left of the middle compartment fell through funnels which Galton referred to as "inclined shoots", that were inclined towards the center of the device. The further from the center a compartment was located, the more was the shoot inclined towards the center. As a result, the original dispersion of the parent population of balls was compressed before the balls were released through the trap doors, at which point the process of family variability (falling down through the rows of pins) increased this dispersion and the offspring population purportedly ended up showing the same dispersion as that of the parent population.[4]

Galton presented a pea-breeding experiment conducted by him with the aid of some friends, which he took to demonstrate the existence of the processes of family variability and reversion (Galton, 1877, pp. 290–1). Galton weighed several thousand pea seeds and determined that the distribution of weight was bell-shaped. He then selected sets of seeds for planting. Each set consisted of seven packets with ten seeds in each packet. Each packet contained seeds from a different section of the distribution. The first packet contained huge seeds whose values belonged to the

[4] The process of reversion described by Galton in 1877 differed to some extent from the original idea of reversion, discussed for instance by animal breeders as well as by Darwin himself, which referred to the return of ancestral characters in a biological population or species (see e.g. Gayon, 1998).

extreme end of the distribution, the seventh packet contained tiny seeds whose values belonged to the other extreme end, and the rest of the packets contained seeds from intermediate sections. Galton prepared seven such sets of seeds and gave them to friends who planted them in several locations around Britain, and measured the sizes of the seeds in the next generation. Galton did not provide any quantitative results of this breeding experiment in his lecture. He wrote that the seeds in each packet produced offspring seeds with weights that distributed in a bell shaped fashion, thus demonstrating according to Galton the process of family variability. The offspring seeds distributed however around a weight value which was not the original value of the parents, but rather around a value closer to the mean weight of the original population as a whole, demonstrating according to Galton the process of reversion.

As Galton himself later stated, "It was anthropological evidence that I desired, caring only for the seeds as means of throwing light on heredity in man" (Galton, 1886b, p. 247), and he went on to demonstrate the existence of reversion in human inheritance. In 1884 he began collecting data on heights of families. He publicly called for families to send him records that included the heights of family members, and promised to reward those who sent appropriate records by paying them out of his own pocket (Galton, 1889b, pp. 72–4). The data he received included the heights of 205 parental couples and 930 of their adult children (Galton, 1886a, p. 52). His analysis of the data he collected on human stature (which later included a second set of data on heights of brothers, see below) appeared in several lectures and papers (Galton, 1885, 1886a, 1886b) as well as in his book *Natural Inheritance* (Galton, 1889b). In contrast with Galton's 1877 work, the quincunx played a minor, if any, role in this analysis. The analysis also included a new term, regression, which replaced the 1877's term reversion, though Galton never stated the reason for this terminological change.[5]

Galton wanted to compare the heights of parents and offspring, but he faced two complications that were absent in his experiments on peas. First, the heights of males and females were not directly comparable because the average height of females was lower than that of males. Galton solved this problem by transforming the female height values by multiplying them by 1.08, a technique that he claimed to have borrowed from fellow anthropologists (Galton, 1886b, p. 247). The second problem was that Galton wanted to investigate the relation between the trait of a parent and that of their offspring, but in contrast to peas, human offspring had two parents rather than one. Galton's solution to this problem was to use the "mid-parent" value, the average height of the pair of parents (Galton, 1886b, pp. 250–1). We do not know if the idea of using the mid-parent value was his own innovation.[6]

Galton's analysis of his data on heights of mid-parents and offspring revealed the following pattern. The height of offspring was on average closer to the median height in the population than was the height of their mid-parents. For instance, there were 11 mid-parents who were 71.5 inches tall in Galton's data, they had 43 offspring overall, and the median height of their offspring was 69.9 inches, which is closer to the median height in the overall population, 68.25 inches, than is 71.5 inches. Galton referred to this pattern, in which offspring traits deviated from the median trait value in the overall population less than their parents' traits, as regression towards mediocrity (Galton, 1886b, p. 246). He determined the value of regression towards mediocrity from mid-parents to offspring to be 2/3 (Galton, 1886b, p. 252; 1889b, p. 95–7). That is, offspring heights tended to deviate from the median height in the overall population only 2/3 as much as their mid-parents did. Galton also determined that mid-parent heights tended to deviate from the median height in the population 1/3 as much as their offspring did (Galton, 1886b, pp. 253–4; 1889b, p. 99–100).

Galton also collected a second set of data. By asking "trusted correspondents" to send him records that included the heights of adult brothers, he received data on 783 brothers from 295 families (Galton, 1886a, p. 52). When he conducted the same kind of analysis that he did on heights of parents and offspring, he found the regression between brothers to be 2/3 (Galton, 1886a, pp. 55–6; 1889b; 108–9).

There is a wide consensus among historians that Galton's attempts to describe inheritance by using statistical tools, which led to the development of the statistical concept of regression, were a significant landmark in the historical development of statistics. Galton's use of the law of deviation and the bell shaped distribution was conceptually different from their traditional use. Traditionally, the bell shaped distribution was conceived to be a result of the effects of accidental causes acting indifferently in any direction and creating a distribution of values centered around a mean value which represented some constant cause, as in Galton's early use of the quincunx (Hilts, 1973, p. 207; 210; 216–7). Thus conceived, the focus was on the mean value at the center of the distribution, because this was the only value with a "real" meaning, describing the effect of a constant cause in the examined population (Hilts, 1973, p. 217). Galton's use of the law of deviation broke away from this tradition because he was not interested in finding constant causes. Rather, he sought a statistical tool to describe the differences between individuals in a population and the way these differences were inherited, and he achieved this goal by describing individuals' deviations from the mean value of the population and the relations between the deviations of parents and those of their offspring. Thus, for Galton the causes that produced the bell shaped distribution were no longer accidental causes, but rather the causal properties of inheritance, and this led him to focus on the entire distribution and not solely on its mean value (Hilts, 1973, pp. 223–4; Mackenzie, 1981, p. 58; Ariew et al., 2015, pp. 643–5; Pence, 2015, pp. 52–3). However, we will argue later that Galton's thoughts about eugenics and human types demonstrate that he did not completely shed the traditional commitment to constant causes.

Galton's conceptual breakthroughs described above enabled the development of statistical tools that would be fruitfully applied to study problems in the life and social sciences, where previous applications of traditional statistical methods mostly proved to be unsuccessful (Hilts, 1973, p. 207; Stigler, 1986, pp. 265, 358–61). In 1888, after *Natural Inheritance* was already sent to press, Galton realized that the analysis he conducted to measure regression between parents and offspring could also be used to measure the relation between the length of a bone and the height of the man it belonged to (a relation which was a key concern for anthropologists at the time), as well as the relation between the anthropometric measures used in Alphonse Bertillon's system of criminal identification (Galton, 1890, pp. 419–21; Porter, 1986, pp. 291–3). The measure he developed to express these relations was the statistical notion of correlation (Galton, 1888).

The next decade would see further substantial development of the concepts of regression and correlation, with Karl Pearson's development of his correlation coefficient measure and linear regression (Stigler, 1986, pp. 342–5), and Udny Yule's formulation of regression as the method of least squares, and of multiple

[5] A few explanations for this terminological change were offered by historians (e.g. Cowan, 1972, p. 524; Porter, 1986, p. 289).

[6] It is perhaps worth noting that even this simple averaging introduces assumptions about inheritance, in particular that both parents make contributions and that they are of equal import. Galton discussed such assumption in his work on eminence, see for example (Galton, 1865).

regression (Stigler, 1986, pp. 349–51, 354–5). These new statistical developments opened the way to new applications of statistics in domains as diverse as evolution and heredity (e.g., with the founding of the biometric school) and economics and public policy (e.g., with studying the effects of the poor laws in Britain). These conceptual breakthroughs in Galton's work and their new applications had led the historian Stephen Stigler to describe Galton's development of the concept of regression as the beginning of a period which he termed "the statistical enlightenment" (Stigler, 2010, p. 469). The philosopher Ian Hacking argued that in Galton's work on regression, chance had finally been "tamed", because for the first time statistical laws were being used for explaining phenomena rather than merely predicting them, and hence they became autonomous in Hacking's terminology (Hacking, 1990, p. 186).[7]

While we agree with the broad consensus among historians regarding the importance of Galton's work, in what follows we will challenge a historical narrative that emerged in the literature and which ascribes an additional statistical breakthrough to Galton. On this historical narrative, after conducting his work on human stature, Galton articulated the modern statistical concept of regression to the mean.

2.3. Galton and the modern statistical explanation of regression to the mean

How did Galton explain the phenomenon of regression towards mediocrity, the fact that the traits of offspring deviated less from the median value compared to their parents? According to what has become a standard narrative, after conducting his work on human stature, Galton came up with a new way of explaining regression towards mediocrity. In lieu of explaining it as being a result of a process of inheritance akin to what he presented in 1877 as reversion, Galton supposedly realized that the pattern was a statistical phenomenon that is now known as "regression to the mean". Using current terminology, the explanation for regression to the mean is as follows. If two variables, x and y, are distributed normally, and there is an imperfect linear correlation between them (i.e., the absolute value of the Pearson correlation coefficient r is less than 1.0), then for each value of x, our best prediction of y will be a value that deviates less from the mean of the distribution of y then the deviation of x from the mean of the x distribution (all measured in standard deviations), and vice versa (Nesselroade et al., 1980, p. 626; Stigler, 1999, pp. 175–6, 187–8).

Describing Galton's work, the historian of statistics Stephen Stigler argued that "it is fair to say that by 1889 Francis Galton had a clear understanding of the concept of regression" (Stigler, 1999, p. 183), and that "Galton had discovered that regression toward the mean was not the result of biological change, but rather was a simple consequence of the imperfect correlation between parents and offspring" (Stigler, 2016, p. 129). More recently, Joeri Witteveen argued that "Galton's 'simple explanation' for regression" appealed to "the *imperfect correlation* between parental and offspring values" (Witteveen, 2019, pp. 13–4). Judea Pearl and Dana Mackenzie write, regarding the statistical explanation of regression to the mean: "By 1889, Galton had figured this out" (Pearl & Mackenzie, 2018, p. 58). In what seems to be a similar interpretation, Ian Hacking concluded that Galton "saw that reversion towards mediocrity was a consequence of the Normal curve" (Hacking, 1990, p. 186). Other historians who wrote on Galton's work did not make this historical argument explicitly, but nevertheless did not distinguish between Galton's concept of regression towards mediocrity and the modern statistical concept of regression towards the mean.

This interpretation of Galton's work has important consequences not only for understanding the development of his statistical thinking, but also for understanding the development of his ideas about the processes of heredity. Recall that in his 1877 talk, Galton described the process of reversion which led offspring to deviate less from the mean than their parents using the inclined "shoots" in his quincunx. According to Stigler, Witteveen and Pearl and Mackenzie, after conducting his work on human stature, Galton stopped appealing to a process of inheritance akin to what he presented in 1877 as reversion and instead explained regression towards mediocrity as a statistical phenomenon (Pearl & Mackenzie, 2018, p. 63; Stigler, 2016, p. 129; Witteveen, 2019, p. 13). Their argument echoes a modern understanding of regression to the mean. Regression to the mean is often discussed today as a phenomenon which prevents one from making an inference from the existence of a pattern of regression to the mean in the data to the existence of a process of regression which causes the value of each y to deviate less than its x counterpart (Nesselroade et al., 1980; Stigler, 1999, pp. 184–6). This blocks inferring a process of reversion from Galton's data on heights. The inference is barred because a pattern of regression to the mean is a necessary consequence of having imperfectly correlated normal distributions; it is a statistical artifact or a statistical effect of the distributions of the data. Stigler, Witteveen and Pearl and Mackenzie seem to imply that Galton understood this, and hence gave up on the idea of an inheritance process akin to reversion.

The narrative we are critiquing makes two historical claims. First, that after conducting his work on human stature Galton explained regression towards mediocrity as a statistical phenomenon, a necessary consequence of other statistical properties of the population. Second, that Galton's explanation of regression towards mediocrity as a statistical phenomenon led him to abandon the idea of a process of inheritance akin to reversion. We will argue that both claims do not hold up. In the next section, we argue that the evidence given to support the first claim is not persuasive. In the following section, we show that Galton's concept of regression towards mediocrity referred to a process of inheritance akin to what he earlier described as reversion.

3. Galton's analysis of regression

We will now go over the main sources of evidence used to support the claim that by the time he wrote *Natural Inheritance* Galton grasped the statistical explanation of regression to the mean. Both Stigler and Pearl and Mackenzie emphasize one graphic method used by Galton to represent regression in his original papers on regression as well in *Natural Inheritance*, which supposedly shows that Galton explained regression towards mediocrity statistically (Stigler, 2016, pp. 126–8; Pearl & Mackenzie, 2018, pp. 60–62). It is instructive to consider in detail Galton's work process leading to this graph. Galton entered his records on mid-parent and offspring heights to the diagram shown in Fig. 2, an earlier version of what we would now call a scatter plot (Pearl & Mackenzie, 2018, p. 60). He claimed that he saw an elliptical contour emerging around the part of the diagram which contained the majority of the entries (he had to merge together groups of four adjacent cells in his table, by averaging their counts, in order to see this pattern better) (Galton, 1886a, pp. 56–7). Galton also noticed that the expected height values of offspring given their mid-parents heights (i.e., the median value of the offspring of mid-parents with a certain value in the data) and vice versa fell on lines, which he called the

[7] There is a growing literature which addresses Hacking's claim regarding the autonomy of statistical laws in Galton's work (Ariew et al, 2015, 2017; Pence, 2015; Witteveen, 2019). We do not discuss this topic in the present paper.

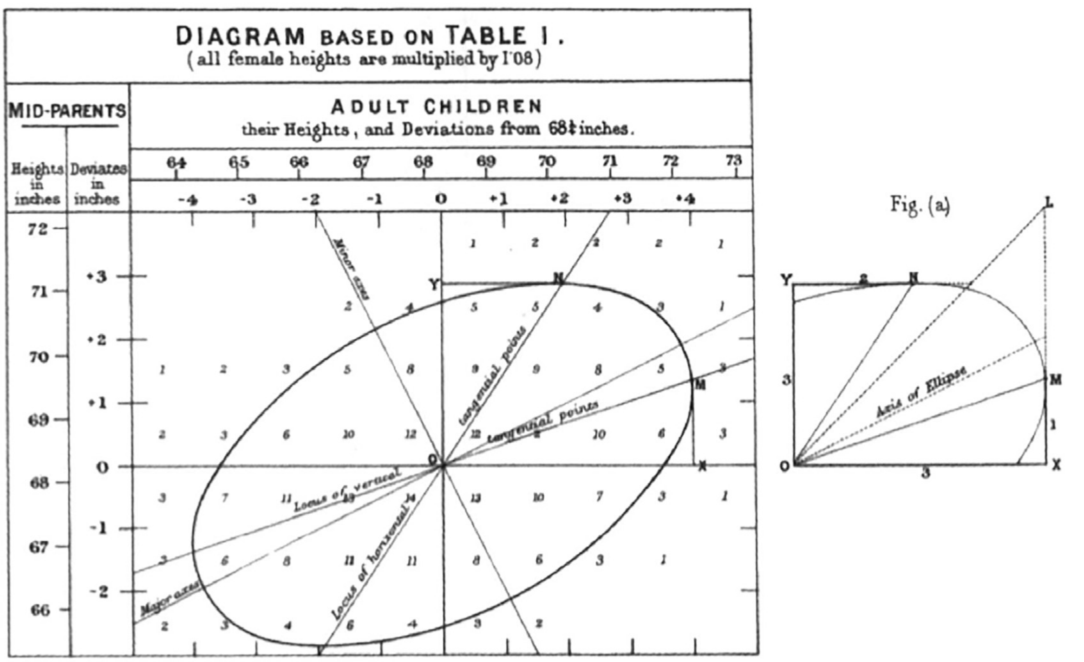

Fig. 2. Galton's diagram presenting regression towards mediocrity in his data (Galton, 1886a).

regression lines, that are less steep than the major axis of the ellipse.

Given Galton's relatively modest mathematical skills, the mathematical analysis in his 1886 paper read to the Royal Society was provided in an appendix by J. D. Hamilton Dickson, Fellow and Tutor of St. Peter's College, Cambridge. The appendix is presented as a purely theoretical exercise and begins in the following way: "Problem 1. A point P is capable of moving along a straight line P'OP, making an angle of $\tan^{-1}\frac{2}{3}$ with the axis of y …. To discuss the "surface frequency" of P." (Galton, 1886b, p. 63) It is in this appendix that we find the formula for an ellipse: $\frac{X^2}{a}+\frac{y^2}{b} = constant$. This formula (number 6 in the appendix) is arrived at after several transformations have been applied to the data.

Both Stigler and Pearl and Mackenzie argue that Galton's diagram clearly shows that given normal distributions and imperfect correlation, regression towards mediocrity must follow (Stigler, 2016, pp. 126–8; Pearl & Mackenzie, 2018, pp. 60–62). The elliptical shape that emerged around the data entries is a result of the normal distributions of the two variables. Since the data are so arranged, the two regression lines can either lie on the major axis of the ellipse if each offspring deviates from the average exactly as much as its mid-parent does (i.e., if there is a perfect linear correlation between the two), or otherwise lie on lines that are less steep than the major axis. When the regression lines are less steep than the major axis of the ellipse, the best prediction of offspring height is one that deviates less from the average than its mid-parent height, and vice versa.

Stigler and Pearl and Mackenzie are not arguing that Galton explicitly described regression towards mediocrity as a consequence of imperfect correlation, because Galton's analysis of his data on stature using the diagram preceded his development of the measure of correlation (correlation was introduced in a 1888 paper, and while *Natural Inheritance* was published in 1889, it was sent to press before Galton wrote his 1888 paper on correlation). We take them to be arguing that Galton's analysis of the diagram shows that he explained regression towards mediocrity to be a necessary consequence of statistical properties of the population, which were later understood as the normal distribution of populations and their imperfect correlation. (We shall take the rest of the evidence that are surveyed in this section to be supporting this kind of claim, because all of the evidence is taken from *Natural Inheritance* or earlier works).

Even if Stigler and Pearl and Mackenzie are right that Galton's diagram shows that regression towards mediocrity is a necessary consequence of other statistical properties, there is no evidence that Galton used the diagram to explain regression towards mediocrity in this way. Galton used the diagram for two purposes.[8] First, he used it to graphically represent regression in his data, using the regression slopes, and to measure the value of regression, which he also measured using other methods (Galton, 1886b, p. 255; 1889b, p. 101–2). Second, he used the diagram to solve an apparent paradox in his data: how could it be that the measure of regression from offspring to mid-parent is not the reciprocal of the measure of regression from mid-parent to offspring (i.e., given that the measure of regression from mid-parent to offspring was 2/3, why was the measure of regression from offspring to mid-parent 1/3 and not 3/2)? (Galton, 1886a, pp. 56–7).[9]

To solve this puzzle, Galton used the ellipse diagram to show the interdependence of the values of four different measures: the quartile of the overall population in his data (his measure of variability), the quartile of the mid-parent population, the regression from mid-parent to offspring, and the regression from offspring to mid-parent.[10] He showed that by knowing any three of these values, we can deduce the fourth value. Using this analysis, Galton showed mathematically that the value of regression from x to y (e.g., from parent to offspring) is not the reciprocal of the value of

[8] Mackenzie (1981, p. 64–5) provides an analysis of Galton's use of the diagram which is similar to ours.

[9] Galton used the obsolete term "converse value" (Galton, 1886a, p. 56).

[10] The equation given by Galton was $c^2w = p^2w'$, in which c is the quartile of the mid-parent population, p is the quartile of the overall population, w is the ratio of regression from mid-parent to offspring, and w' is the ratio of regression from offspring to mid-parents (Galton, 1886a, p. 57).

regression from y to x (from offspring to parent) and dissolved what initially seemed to him to be a paradox. Showing why the measure of regression between parents and offspring was not the reciprocal of the measure of regression between offspring and parents was not used by Galton to explain why offspring deviate less than their parents and vice versa.

Witteveen highlights Galton's use of the law of ancestral inheritance to give a different argument purportedly showing that Galton explained regression towards mediocrity statistically (Witteveen, 2019, pp. 12–4). The law of ancestral inheritance was developed by Galton himself. According to the law, each individual inherits part of their traits from their parents and part of their traits from more ancient ancestry (Galton, 1885, p. 1209).[11] The further we go back in the individual's ancestry, the more it will represent the overall population with its average traits. Thus, offspring to parents who deviate significantly from the mean inherit both from their parents and from remote ancestors who are mediocre, or have average traits, and as a result will deviate less than their parents (Galton, 1885, p. 1209). According to Witteveen, the essence of Galton's explanation here was that given imperfect correlation, regression towards mediocrity must follow (Witteveen, 2019, pp. 13–4). Galton described the processes by which offspring inherited both from their parents and from more remote ancestors to make the point that offspring only partly resembled their parents and hence that there was an imperfect correlation between the two. This imperfect correlation in turn implied that the ratio of regression must take some value between 0 and 1. Thus, Galton's explanation of regression did not explicitly appeal to the specific details of the processes of inheritance but only to the fact that these processes resulted with partial resemblance, or imperfect correlation, between parents and offspring (Witteveen, 2019, p. 13–4).

We agree with Witteveen that explaining regression towards mediocrity as the result of partial resemblance between parents and offspring should count as explaining it statistically. And yet, this was not Galton's explanation here. We cannot find any quote in which Galton emphasized the partial resemblance between parents and offspring in his explanation of regression towards mediocrity by appeal to ancestral inheritance. Rather, his explanation rested on the idea that the trait of an individual was the result of combining the hereditary material received from its parent with that received from remote ancestry, and especially on the claim that the remote ancestry was necessarily mediocre (Galton, 1885, p. 1209; 1886b, p. 252–3). Accordingly, if the parent deviated from the mean, the offspring's trait would deviate less because it was a mixture of the deviating parent and the more mediocre ancestry. Galton's explanation was not that regression towards mediocrity was a consequence of partial resemblance between parents and offspring, but rather that it was a consequence of the mixing in the offspring of a parent deviation and the mediocre remote ancestry. We get a glimpse into Galton's thinking in the metaphor he used to illustrate his explanation. He compared the deviating parent to a vessel of pure wine, the mediocre ancestry was compared to some amount of water, and the offspring was compared to the diluted wine produced by pouring the water into the wine vessel (Galton, 1885, p. 1209; 1886b, p. 253).

Stigler gives two quotations from Galton's 1889 book, *Natural Inheritance*, that are supposed to demonstrate Galton's statistical explanation of regression towards mediocrity. First, he provides the following quotation from Galton: "However paradoxical it may appear at first sight, it is theoretically a necessary fact, and one that is clearly confirmed by observation, that the Stature of the adult offspring must on the whole, be more *mediocre* than the stature of their Parents" (Galton, 1889b, p. 95; quoted in; Stigler, 1999, p. 183). This quote can be understood as claiming that regression toward mediocrity is theoretically necessary because statistical theory shows that it is a consequence of other statistical properties. And yet Galton refers to regression towards mediocrity both as a theoretically necessary fact and as a fact that can be observed in the data. The rest of the sub-section which begins with this quote does not deal with statistical theory at all, but rather with the demonstration of regression towards mediocrity in Galton's data (Galton, 1889b, pp. 95–9). As such, it is not clear which theory exactly made regression a necessary fact according to Galton; whether statistical theory or a theory regarding the physiological processes of inheritance. The fact that the first two chapters of *Natural Inheritance* (following an introductory chapter) are entirely dedicated to a discussion of the processes of inheritance and to the stability of types favors the second option.

Second, Stigler (2016, p. 128) points to a quote from *Natural Inheritance* in which Galton explained the pattern of regression in height between brothers: " … the unknown brother has two different tendencies, the one to resemble the known man, and the other to resemble his race. The one tendency is to deviate from P [the mean stature] as much as his brother, and the other tendency is not to deviate at all. The result is a compromise" (Galton, 1889b, p. 109). If anything, this quote seems to suggest the existence of two causes or processes, each pushing the individual's stature in a different direction, with the resulting height being a compromise.

Stigler and Witteveen also argue that the fact that Galton identified regression towards mediocrity not only between parents and offspring but also between brothers and between offspring and parents shows that he understood it to be a statistical phenomenon (Stigler, 2016, pp. 124–6; Witteveen, 2019, p. 14). They argue that since Galton could not have possibly thought that the trait of a brother caused the trait of his brother to deviate less from the mean or that traits of offspring caused their parents to deviate less, he must have realized that the explanation for regression towards mediocrity in these cases is statistical. This claim, that Galton had no other explanation to appeal to other than the statistical explanation, assumes from the get-go that this statistical explanation was available to Galton. Even more crucially, it is incorrect that the only causal explanation available to Galton was one in which the trait of the brother or the offspring caused the trait of the other brother or the parent to revert. Instead, Galton could have appealed to a common cause that affected both brothers or both offspring and parents, resulting in the pattern of regression towards mediocrity. In the next section, we will argue that the quote given in the previous paragraph most likely referred to such an explanation.

Stigler, Pearl and Mackenzie and Witteveen do not discuss Galton's works on correlation, a measure he developed in 1888, after he wrote the works they focus on. Galton's work on correlation is quite telling. He noted how the same analysis that enabled him to measure regression between relatives was used to measure correlation between body parts (Galton, 1890, p. 421). He also gave the mathematical relation between the measures of correlation and regression (Galton, 1890, p. 427). And yet, we cannot find in Galton's discussion of correlation any hint that he sought to explain regression towards mediocrity as a consequence of imperfect correlation. In fact, the opposite might be the case. In describing the mathematical relation between regression and correlation, Galton wrote that "If there is no regression at all … then the correlation becomes identity. If the regression is complete … there is no resemblance at all" (Galton, 1890, p. 427) and also referred to the "completeness with which the ratio of regression measures correlation" (Galton, 1890, p. 429). These statements may suggest that

[11] In quantitative terms, the law of ancestral inheritance said that each child inherits 1/2 from its parents, 1/4 from its grand-parents, 1/8 from great grandparents, and so on (Galton, 1886a, p. 62).

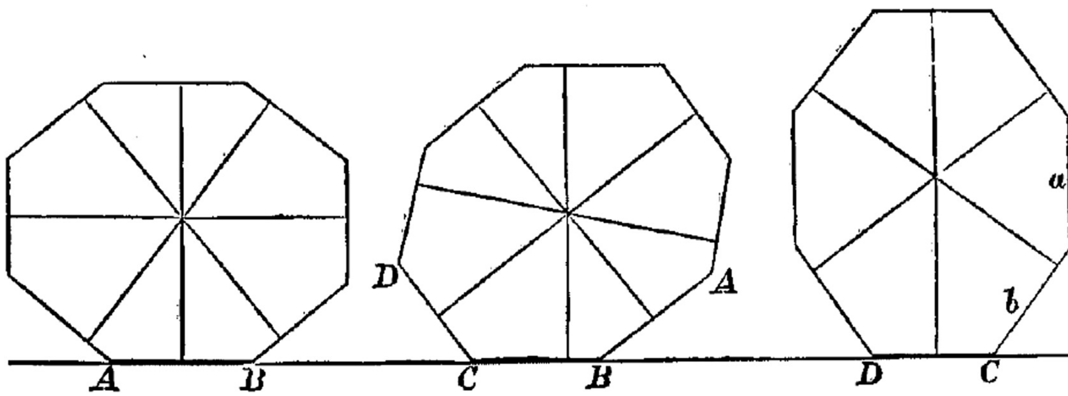

Fig. 3. Galton's polygon model (Galton, 1889).

for Galton the measure of regression between two populations explained the measure of their correlation. However, the evidence is too scarce because, in contrast with regression, Galton wrote very little on correlation, and shortly after developing this statistical notion he moved on to other endeavors (Cowan, 1972, p. 527).

We conclude that the evidence given to support the claim that Galton explained regression towards mediocrity as a statistical phenomenon is not persuasive. We now turn to challenge the claim that upon conceiving of regression as a statistical phenomenon, Galton abandoned the idea of a process akin to what he described in 1877 as reversion.

4. Galton's speculations about the processes of inheritance

In 1877, Galton identified a process of reversion illustrated by the inclined "shoots" in his quincunx. An examination of Galton's ideas regarding inheritance and evolution after 1877 shows that he continued to appeal to such a process in order to explain regression towards mediocrity between parents and offspring throughout his work. This gives additional support to the conclusion that Galton did not explain regression towards mediocrity statistically.

Galton came up with two explanations for the pattern of regression towards mediocrity. The first was his idea of ancestral inheritance, the fact that each individual inherits partly from his parents and partly from more remote ancestors. The second explanation was the organic stability of types. Galton's ideas of stability of types go back at least to his book *Hereditary Genius*, where a short discussion of stability can be found (Galton, 1869, pp. 368–70). Over the years, Galton connected the ideas of regression and organic stability and came to regard stability as an important explanation for regression towards mediocrity between parents and offspring.

In one of the original papers presenting his analysis of human stature and the phenomenon of regression towards mediocrity between parents and offspring, Galton explained regression towards mediocrity by appeal to ancestral inheritance, but he also argued that "The stability of a type … would … be measured by the strength of its tendency to regress" (Galton, 1886b, p. 258), thus connecting regression with the notion of stability. In *Natural Inheritance*, Galton explicitly referred to both ancestral inheritance and organic stability as possible explanations for regression towards mediocrity between parents and offspring (Galton, 1889b, pp. 104–6). Galton dedicated the entire third chapter of *Natural Inheritance* to a discussion of the organic stability of types, indicating the importance he assigned to it. Five years later, in a review of William Bateson's book *Materials for the Study of Variation*, Galton argued that stability was the most important cause of regression towards mediocrity between parents and offspring (Galton, 1894, p. 364).

Galton's explanation of organic stability began with the observation that the processes by which offspring came to resemble their parents (which we would now refer to as the processes of inheritance and development), were full of disturbances (Galton, 1889b, pp. 20–1). Some variants of a trait, for instance certain heights, were more stable than others, and were thus referred to by Galton as types.[12] As a result, the stable variants, belonging to a type, would leave offspring that closely resemble them, since the stability of the variant is more resistant to disturbances in inheritance and development. In contrast, the disturbances would cause offspring of the less stable variants, that deviate from the type, to revert back to the more stable type (Galton, 1889b, p. 104). Thus, on this conception of stability, a bell shaped distribution of a trait in a population was the result of a stable type together with disturbances (Galton, 1894, p. 364).[13] The mean value of the trait, which was also the most frequent value in the population, was the type, and variants that deviated from the mean were unstable variants. The instability of these variants caused their offspring to regress back to the type.

Galton's ideas about stability of types were inspired by mechanical stability. He presented a model consisting of the stable resting positions of a polygon shaped object as shown in Fig. 3 (Galton, 1889b, pp. 27–30) (an earlier version of the polygon and its analogy to stability already appeared in Galton's book *Hereditary Genius*, where the polygon was described as a rough stone (Galton, 1869, p. 369)). The figure shows a solid polygon in three different orientations. The polygon's orientation can be changed by tilting (thereby representing disturbances). When the polygon rests on edge AB (displayed on the left side of the figure), it will be difficult to move it to the orientation to the right in which it rests on edge BC, and will require intense tilting. When it rests on edge BC, it will be moreover much less stable, since it is highly probable that once it is tilted again it will fall back to rest on edge AB. Hence, the orientation in which the polygon rests on AB is a much more stable orientation than the one in which it rests on BC.

Though Galton did not explicitly appeal to his ideas about organic stability to account for regression towards mediocrity between brothers, we find it probable that this is what he had in mind when he explained this regression pattern to be the result of

[12] Galton also provided a hierarchy of types and sub-types (Galton, 1889b, pp. 25–7), but this point shall not concern us here.

[13] Galton sometimes referred to the type and the distribution around it as a race (e.g. Galton, 1894, pp. 362–3).

a compromise between two tendencies: that of a man to resemble his brother and that of a man to resemble his race (Galton, 1889b, p. 109). A man's tendency to resemble his brother is a result of inheriting similar traits from the same parents. A man's tendency to resemble the race is a result of organic stability. A deviating trait, while caused by the shared hereditary endowment of two brothers, is unstable, and the probability that the trait will appear in both of the brothers is low. The more probable result would be a trait that is a compromise between the stable and deviating trait and hence closer to the population mean. The same analysis could be applied to regression towards mediocrity between offspring and parents.

The period during which Galton came to regard organic stability as the most important cause of regression towards mediocrity saw additional development in his ideas about regression and possible change in biological populations. Galton took regression towards mediocrity in his data to be the result of an inheritance process that kept the distributions of traits in populations stable across generations, in a similar fashion to the way reversion was taken to play this role in his earlier writings. Already in 1877 Galton argued that population stability was a result of the counteracting effects of the process of family variability and the process of reversion, which he described using the quincunx. He offered the same explanation for population stability in a paper presenting his work on human stature and regression towards mediocrity in 1886 (Galton, 1886b, p. 256), as well as in *Natural Inheritance* (Galton, 1889b, pp. 116–7). There were two counteracting processes of inheritance, one which increased the dispersion of the trait, and another one which led offspring to regress closer to the mean and hence decreased the dispersion of the trait. The result was a constant distribution of the trait across generations. In both analyses, his measure of the process which led offspring to regress closer to the mean was the measure of regression towards mediocrity obtained from his data on stature.

Shortly after Galton finished writing *Natural Inheritance*, he began to strongly promote the argument that regression also prevented gradual evolutionary change by natural selection. We can already find the roots of this idea in *Natural Inheritance* (Galton, 1889b, pp. 32–3), as well as 20 years earlier in *Hereditary Genius* (Galton, 1869, p. 369), but after writing *Natural Inheritance* Galton had developed it to the firm conclusion that gradual evolution was impossible. This argument appeared in Galton's Presidential Address to the Anniversary Meeting of the Anthropological Institute of Great Britain and Ireland in January 1889 (Galton, 1889a, p. 406), in his book *Finger Prints* (Galton, 1892a, pp. 211–20), in the second edition of *Hereditary Genius* (Galton, 1892b, p. xvii-xix) and in his 1894 review of William Bateson's *Materials for the Study of Variation* (Galton, 1894, pp. 363–4, 368).

In his 1894 review, Galton asked the reader to imagine a population of individuals dispersed around trait A, which changed after a long course of evolution to a population which is dispersed around trait B. How could the evolution of this population proceed from A to B? Darwin, Galton noted, stressed that evolution proceeded by minute steps. This meant that small variants which slightly differed from A and were closer to B would be favored by natural selection, leaving more offspring that resembled them. Some of these offspring would differ more from A than others and would be slightly closer to the value of B, and these again would be favored by selection. This process would go on until finally the population would consist of individuals dispersed around trait B.

But, Galton countered, this mode of evolution by minute steps disagreed with what was known about heredity: "There can be no doubt as to the reality of regression. I have not only proved its existence in certain cases and measured its amount, but have also shown that no race could continue constant in its characteristics unless regression existed. And, again the observed and the theoretical details of the process were found to strictly concur" (Galton, 1894, p. 364). Accordingly, if a population was dispersed around the trait value of A, this meant that A is a stable type, and heredity would regress to it. The conclusion Galton reached was that evolution could not proceed by minute steps, because even if variants that deviated slightly from A were selected for and hence left more offspring, the offspring of these variants "would tend to regress backwards towards the typical center, and the advance that had been made would be temporary and could not be maintained" (Galton, 1894, p. 368). Rather, evolution could only proceed by the sudden creation of "sport" variants that were qualitatively different from their parents and constituted new stable types of a trait. In Galton's own words: "by a leap from one position of organic stability to another" (Galton, 1894, p. 368).[14] This analysis had direct implications for eugenics.

Galton's ideas of organic stability, his explanation of population stability being the result of opposing inheritance processes, and his arguments regarding the impossibility of gradual evolution are important to our discussion for several reasons. These ideas show that in contrast to the claim made by Stigler, Witteveen and Pearl and Mackenzie, Galton never gave up on the idea of an inheritance process akin to what he described in 1877 as reversion. Rather, Galton explained regression towards mediocrity between parents and offspring to be the result of such a process of inheritance. Galton argued that regression towards mediocrity was the result of the instability of the variants that deviated from the mean, which caused their offspring to deviate less. This was a specific physiological process that could underlie the reversion phenomenon described in 1877, in which offspring of parents who deviated from the mean reverted back to the mean value due to the inclined "shoots". Galton also took regression towards mediocrity as a measure of a process that counteracted the increase in variability of traits and hence kept populations stable, just like reversion operated in his 1877 lecture, and that prevented gradual evolutionary change by natural selection. In other words, for Galton regression towards mediocrity was a result of the processes of inheritance that took place in biological populations and affected how these populations could change. Had Galton explained the pattern of regression towards mediocrity merely as a statistical phenomenon by this time, he would have had no reason to conclude that regression kept populations constant and prevented gradual evolutionary change. The statistical phenomenon is a consequence of the distribution of the parent and offspring populations and cannot tell us how future change will occur.

Galton's Presidential Address to the Anniversary Meeting of the Anthropological Institute of Great Britain and Ireland in January 1889 is instructive. The talk marked the conclusion of his four-year presidency of the institute. Galton devoted a large part of the talk to a discussion of correlation, noting that it existed "wherever the variations of two objects are in part due to common causes" (Galton, 1889a, p. 403). He pointed out that "family resemblance" (not *reversion*) was a special case of correlation, and used correlation to explain the originally puzzling case of the heights of brothers. He also extolled the benefits of finding order by looking at the distribution of traits at the population level over the disorder of individual differences (noting, in passim, that "the average worth of the Hebrew race" is low, but that the race displays great variability).

[14] It is worth noting that Galton's position on evolutionary jumps were described as ambiguous by some historians (e.g. Bowler, 2014). Nevertheless, we take his works after 1890 to be clearly advocating the idea that evolution proceeded in jumps and not in minute steps. We thank an anonymous referee for commenting on this point.

But, in what appears as a throwback to his earlier ideas about differently inclined "shoots", that is about a process differentially affecting parents in different parts of the distribution, Galton went on to claim that the average man was a source of evolutionary inertia and was by nature "essentially unprogressive" because "his children tend to resemble him exactly, whereas children of exceptional persons tend to regress to mediocrity" (Galton, 1889a, p. 406).

In contrast with the common historical narrative which focused on the development of statistics and asked how Galton explained regression towards mediocrity in his data, we saw that explaining this pattern was part of Galton's endeavor to answer additional biological questions about biological types and population stability across generations. At least some of these questions could only be solved by appealing to processes of inheritance. In particular, the problem of population stability required a description of a process or several processes of inheritance that made the distribution of traits constant over generations (Krashniak & Lamm, 2017). The problems that concerned Galton the most called for understanding the processes of inheritance.

Galton's ideas about the relation between regression and types also show that he did not completely abandon the statistical tradition which focused on the mean value of the bell shaped distribution and conceived of it as reflecting a constant cause. While Galton rejected the idea that the distribution around the mean reflected accidental causes which were of little importance, he kept ascribing special importance to the mean value at the center of the distribution, taking it to reflect a stable type that variations regressed to (see also Porter, 1986, pp. 139–40; Cryle & Stephens, 2017, p. 224).

5. Regression to the mean in the biometrician-mutationist debate

While Galton did not explain regression towards mediocrity statistically and did not as a result give up on the idea of an inheritance process of reversion, the statistical explanation of regression to the mean was indeed used to reject the idea of a reversion/regression inheritance process which led offspring to revert to the mean. This development however took place in the early 1900's and was made by Galton's successors. It was part of the debate about evolution between the biometric camp (which included Raphael Weldon and Karl Pearson) and the mutationist-mendelian camp (which included William Bateson and Hugo De Vries). Galton's concept of regression towards mediocrity played a different role in this development than the one ascribed to him in the usual historical narrative.

The debate between the biometricians and the mutationist-mendelians has received considerable scholarly attention (e.g., Provine, 1971; Olby, 1989; Gillham, 2001; Radick, 2011; (Radick,); Pence, forthcoming; Shan, 2020). The origins of this debate can be traced to a 1895 debate between Weldon and Bateson over the question of whether the variation that supplied the material for evolution was continuous or discontinuous (Provine, 1971, pp. 44–6). Karl Pearson soon joined Weldon's side, and the debate took off and heated in the years 1900–1906, cooling down after Weldon's sudden death in 1906. The debate revolved around several points of controversy regarding inheritance and evolution (see Olby, 1989, p. 300), but the point of controversy that shall concern us here is the question of whether evolution proceeded gradually, by selection of small variations, or in big evolutionary jumps. The biometricians argued that evolution was gradual, while the mutationists argued that evolution took place in jumps.

Both sides in the debate were influenced by Galton and claimed to be his true successors. Pearson and Weldon adopted Galton's law of ancestral inheritance, as well as his statistical tools of regression and correlation, and further developed them in order to study inheritance (Provine, 1971, ch. 2). Bateson and De Vries, in turn, were influenced by Galton's ideas regarding evolution in jumps (Gillham, 2001, pp. 288–9; Gould, 2002, p. 344). In what follows we will highlight the way in which the concept of regression to the mean became a key point of contention between the mutationists and the biometricians.

Bateson's argument that new biological traits and species most likely appeared as a result of big changes rather than accumulation of small changes by natural selection goes back at least to his 1894 book *Materials for the Study of Variation* (Bateson, 1894) (Galton highly praised the book in his review which we discussed earlier). His argument for saltational evolution was sometimes based on ideas about regression and types which closely resembled Galton's. Bateson distinguished between types and varieties. He argued that a type comprised a bell shaped distribution, with the mean value at the center, which was also the most frequent value in the distribution, representing the value of the type (Bateson, 1897, p. 555). The bell shaped distribution itself represented what Bateson referred to as the varieties of the type. Thus, if a population consisted entirely of varieties of one type, it would show a bell shaped distribution with one peak, while if it were a mixture of two types it would show a distribution with two peaks, and so on (Bateson, 1897, p. 555). Offspring of varieties tended to regress to the mean or the center of the distribution of the type that they belonged to (Bateson, 1897, pp. 555, 557; Bateson, 1899, p. 4). One conclusion that Bateson seemed to extract from this discussion of regression was that evolution could not proceed by the accumulation of small changes, or varieties, because these varieties will regress back to their type (Bateson, 1899, pp. 3, 4). Hence, evolution could only proceed by the appearance of new types. As we showed earlier, this was essentially Galton's analysis.

In his two-volume book, *The Mutation Theory* (published in German in two parts, in 1901 and 1903, and translated to English in 1909), De Vries argued that the phenomenon which Galton described as regression prevented the creation of new species and traits by the selective accumulation of small changes. De Vries demonstrated the occurrence of regression using the same methodology that Galton used. For instance, he planted seeds of *Madia elegans*, which varied in the ray-floret on the head of the flowers, and showed that the ray-floret number of offspring plants deviated only about as 1/3 from the mean number as their parents did (De Vries, 1909, pp. 83–4). Regression, De Vries argued, prevented evolutionary change of small variations by natural selection, because these variations would regress back to the mean and the population's distribution would remain constant (De Vries, 1909, pp. 120, 161–2). De Vries distinguished between small, or regular, variations, and mutations (De Vries, 1909, pp. 135–6). While deviating variations regressed back to the old mean, mutations were big qualitative changes that had a new mean towards which they regressed. Evolution, he concluded, could not proceed by the selection of variations but only by the appearance of new forms due to mutation (De Vries, 1909, pp. 211–2).

Pearson's and Weldon's attacks on Bateson's and De Vries's ideas regarding evolution by jumps involved criticizing their use of the concept of regression. While Bateson and De Vries were not interested in statistics per se and did not have a deep understanding of statistical tools such as Galton's regression, Pearson and Weldon (especially Pearson) made substantial contributions in the 1890's to the further development of Galton's statistical techniques.

In a 1902 paper, Pearson argued that Bateson's ideas consisted of "a hopelessly confused notion of what we are to understand by regression", and that " [t]he concept of regression is equally obscure in Professor Hugo De Vries' ideas" (Pearson, 1902, p. 322).

He argued that in contrast to Bateson's and De Vries's description of regression as a biological phenomenon, statisticians understood that regression was a phenomenon that could be observed whenever there were two varying characters, biological or not (Pearson, 1902, pp. 322–3). While Pearson did not argue explicitly in this paper that regression towards mediocrity was a statistical phenomenon, a consequence of other statistical properties, the argument that regression could take place between any two variables is a clear indication that this is the kind of notion he had in mind.

+ In the same year, Weldon published a paper in which he criticized De Vries's arguments that evolution occurred in jumps and not via gradual change. Weldon argued that De Vries's ideas regarding evolution by jumps were fundamentally based on his conception of regression towards mediocrity (Weldon, 1902, p. 366), and turned to show, like Pearson, that this conception did not conform with the statistical understanding of regression. + Weldon argued that Pearson's work on regression had demonstrated that the focus of regression in each generation is the mean value of *that* generation, and not some fixed value that does not change (Weldon, 1902, p. 368). Thus, a pattern of regression in a population does not imply that there is a process of inheritance that causes offspring to regress towards a fixed value of a specific type, because the mean value towards which regression will be directed can itself change. Again, although not stated explicitly, this argument indicates that Weldon took regression to the mean to be a statistical phenomenon, though he conflated here the regression to the mean that is found between a parental and an offspring population and the question of why each population is normally distributed. We can see in Pearson's and Weldon's papers how the debate was being shaped to be one about understanding the statistical notion of regression rather than about the underlying causes, whether in terms of inheritance or development. Weldon concluded his paper by expressing the hope that a better statistical understanding would lead De Vries and Bateson to give up their conception of a "relation between the phenomenon of "regression" and the stability of specific mean character through a series of generations" and the evolutionary ideas that were based on this conception (Weldon, 1902, p. 374).

In a lecture given in Oxford in 1905, and published in 1906, Weldon explicitly explained regression towards mediocrity statistically, as a consequence of imperfect correlation between two variables. In a beautiful informal demonstration, he illustrated this explanation with a dice tossing experiment (Weldon, 1906, pp. 100–6). Weldon demonstrated regression towards mediocrity between two series of dice tosses, and argued that it was a necessary consequence of the imperfect correlation between these series of tosses. He used this demonstration to criticize those biologists who, for lack of statistical understanding, referred to regression to the mean as "a peculiar property of living things, by virtue of which variations are diminished in intensity during their transmission from parent to child, and the species is kept true to its type" (Weldon, 1906, p. 107).

Pearson and Weldon argued that their use of the concept of regression was similar to Galton's, and that Bateson and De Vries did not understand Galton's ideas about regression to mediocrity and distorted them for their own use (Pearson, 1902, p. 323; Weldon, 1902, p. 369; 1906, p. 106–7). But as we have shown, Galton's concept of regression towards mediocrity was actually very close to that of Bateson and De Vries. Like them, Galton described regression towards mediocrity as a measure of an inheritance process that caused offspring to regress back to some fixed type, and hence kept populations fixed and precluded evolution from proceeding by the accumulation of small changes. In the third volume of his biography of Galton, published in 1930, Pearson acknowledged that Galton did not understand regression towards mediocrity to be a statistical phenomenon, but rather "this idea of regression fixing itself in his mind became for him a biological fact" (Pearson, 1930, p. 76). With his ideas on regression as a process that caused offspring to revert back to the mean and that hindered evolution by small changes, wrote Pearson, Galton "was a mutationist before De Vries published his first paper on mutations" (Pearson, 1930, p. 32).

Thus, while Bateson and De Vries adopted Galton's concept of regression towards mediocrity, Pearson and Weldon developed the modern statistical concept of regression to the mean that we recognize today. The disagreement between the two sides regarding the use of the regression concept was not over a technical issue concerning statistical methodology, it was part of a larger debate on whether evolution proceeded by accumulating small changes or by jumps. + Indeed, one may wonder whether this evolutionary debate with the mutationists is what drove Weldon and Pearson to thoroughly articulate the idea of regression as a statistical phenomenon, though more evidence is needed to support such a claim.

The historical claim that Galton explained regression towards mediocrity as a statistical phenomenon by the time he wrote *Natural Inheritance* misses the much more gradual development of this statistical concept of regression from Galton's concept to Pearson's and Weldon's, the context of the later evolutionary debate in which the modern statistical concept was developed, and the role that Galton's concept of regression to mediocrity played in this debate. imp It privileges a framing that was introduced by the biometricians without identifying it as such. It also suggests more conceptual continuity than we should expect or have evidence for. Rather than seeing a bifurcation, with the biometricians and mutationists each latching on to different aspects of Galton's primordial thoughts and developing different notions, the narrative that was developed mainly in the context of the history of statistics focuses on a single line of development. This line of development highlights aspects of Galton's thought that can be seen as leading to the modern statistical notion of regression to the mean at the risk of missing the biological debate and the concerns and questions motivating the actors.

6. Concluding remarks

In contrast with what has become a common narrative in the history of statistics, according to which Galton articulated the statistical explanation for regression to the mean and hence gave up on the idea of an inheritance process akin to what he previously described as reversion, we showed that Galton did not explain regression towards mediocrity statistically and did not abandon the idea of a process of reversion. The development of the statistical concept of regression to the mean was more gradual than is often portrayed, and took place in a different context. In fact, Galton's concept of regression towards mediocrity was significantly different than the modern statistical concept of regression to the mean and throughout his life he explained regression towards mediocrity by appeal to a process of reversion in inheritance, which caused offspring of parents who deviated significantly from the mean trait to revert back to the mean.

Galton is best viewed as a transitional figure in the understanding of the statistical phenomenon of regression to the mean. On the one hand, he developed statistical measures of regression and correlation and showed that these measures were mathematically related. On the other hand, Galton did not explain regression towards mediocrity statistically, as a consequence of imperfect correlation or any other statistical properties for that matter.

Historians who argued that Galton explained regression towards mediocrity statistically neglected Galton's ideas on inheritance and evolution, specifically his continuing analysis of the inheritance process of reversion, as well as the biological concerns which motivated this work. Focusing on the history of the statistical notion of regression alone detracts attention from two fundamental issues: the role played by concerns about types (in particular, human types) and the explanatory burden, if any, that the statistical explanation has for the biological concerns that motivated the debates. As we saw, the notion of types played a role in anthropology writ large, and in the conception of human groups and races. Galton's work should be seen in this context. However, the biological concerns of Galton, and the rise of evolutionary as well as of developmental thinking, put stress on the appropriate notion of type: as a characteristic of individual development (as for instance illustrated by Galton's polygon model) or as reflected by population level properties (as illustrated by Galton's original use of the quincunx and by Quetelet's notion of *homme moyen*). How these two notions are related and the implications for the mutability or immutability of types was a critical concern. These issues also had direct bearing on conceptions of race and of eugenics. Galton's theoretical work on heredity was driven by such concerns and his work on regression towards mediocrity cannot be seen independently from the question of type. He dedicated chapter 3 of his book *Natural Inheritance* to organic stability, while chapters 4 and 5 were dedicated to population frequencies and the normal distribution. In the book's summary, Galton noted that a complete theory of heredity had to conform to two observations, reflecting two questions that he addressed throughout his career: the particulate nature of inheritance and the stability of the organism (Galton, 1889b, p. 193).

Declaration of competing interest

The authors declare that they have no conflicts of interest.

Acknowledgements

We would like to thank Yafeng Shan, Snait Gissis, Shira Shmuely, Amos Morris-Reich, and the two anonymous referees for their valuable comments on this paper. Adam Krashniak's work was kindly supported by The Interuniversity Ph.D. Program in the History and Philosophy of the Life Sciences, supported by the Humanities Fund of the Israeli Council of Higher Education.

References

Ackerknecht, E. H. (1953). *Rudolf Virchow: Doctor, statesman, anthropologist*. Madison: University of Wisconsin Press.
Ariew, A., Rice, C., & Rohwer, Y. (2015). Autonomous-statistical explanations and natural selection. *The British Journal for the Philosophy of Science, 66*(3), 635-658.
Ariew, A., Rohwer, Y., & Rice, C. (2017). Galton, reversion and the quincunx: The rise of statistical explanation. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 66*, 63-72.
Bateson, W. (1894). *Materials for the study of variation treated with especial regard to discontinuity in the origin of species*. London: Macmillan.
Bateson, W. (1897). On progress in the study of variation. *Science Progress, 6*(5), 554-568, 1894-1898.
Bateson, W. (1899). Hybridisation and cross-breeding as a method of scientific investigation. *Journal of the Royal Horticultural Society, 24*, 59-66.
Boas, F. (1893). *Remarks on the theory of anthropometry* (Vol. 3, pp. 569-575)Publications of the American Statistical Association, 24.
Bowler, P. J. (2014). Francis Galton's saltationism and the ambiguities of selection. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 48*, 272-279.
Cowan, R. S. (1972). Francis Galton's statistical ideas: The influence of eugenics. *Isis, 63*(4), 509-528.
Cryle, P., & Stephens, E. (2017). *Normality: A critical genealogy*. Chicago: University of Chicago Press.
De Vries, H. (1909). *The mutation theory: The origin of varieties by mutation* (Vol. 1) Chicago: Open Court Publishing Company.
Desrosières, A. (1998). *The politics of large numbers: A history of statistical reasoning*. Cambridge: Harvard University Press.
Fancher, R. E. (1983). Francis Galton's african ethnography and its role in the development of his psychology. *The British Journal for the History of Science, 16*(1), 67-79.
Galton, F. (1865). *Hereditary talent and character* (Vol. 12, pp. 318-327)Macmillan's magazine, 157-166.
Galton, F. (1869). *Hereditary Genius: An inquiry into its laws and consequences*. London: Macmillan.
Galton, F. (1877). Typical laws of heredity. *Proc. Of the Royal Institution, 8*, 282-301.
Galton, F. (1885). Presidential address, section H, anthropology. *British Association Reports, 55*, 1206-1214.
Galton, F. (1886a). Family likeness in stature. *Proceedings of the Royal Society of London, 242*(40), 42-73.
Galton, F. (1886b). Regression towards mediocrity in hereditary stature. *Journal of the Anthropological Institute of Great Britain and Ireland, 15*, 246-263.
Galton, F. (1888). Co-relations and their measurement, chiefly from anthropometric data. *Proceedings of the Royal Society of London, 45*(273–279), 135-145.
Galton, F. (1889a). Human variety [address delivered at the annual meeting of the anthropological Institute of great Britain and Ireland]. *Journal of the Anthropological Institute, 18*, 401-419.
Galton, F. (1889b). *Natural inheritance*. London: Macmillan.
Galton, F. (1890). Kinship and correlation. *North American Review, 150*(401), 419-431.
Galton, F. (1892a). *Finger Prints*. London: Macmillan.
Galton, F. (1892b). *Hereditary Genius: An inquiry into its laws and consequences* (2nd ed.). London: Macmillan.
Galton, F. (1894). Discontinuity in evolution. *Mind, 3*(11), 362-372.
Galton, F. (1908). *Memories of my life*. London: Methuen.
Gayon, J. (1998). *Darwinism's struggle for survival: Heredity and the hypothesis of natural selection*. Cambridge: Cambridge University Press.
Gillham, N. W. (2001). *A life of sir Francis Galton: From african exploration to the birth of eugenics*. Oxford: Oxford University Press.
Gould, S. J. (2002). *The structure of evolutionary theory*. Cambridge: Harvard University Press.
Hacking, I. (1990). *The taming of chance*. Cambridge: Cambridge University Press.
Hilts, V. (1973). Statistics and social science. In R. N. Giere, & R. S. Westfall (Eds.), *Foundations of scientific method: The nineteenth century* (pp. 206-233). Bloomington: Indiana University Press.
King, C. (2019). *Gods of the upper air: How a circle of renegade anthropologists reinvented race, sex, and gender in the twentieth century*. New York: Doubleday.
Krashniak, A., & Lamm, E. (2017). Was regression to the mean really the solution to Darwin's problem with heredity? *Biology and Philosophy, 32*(5), 749-758.
Mackenzie, D. A. (1981). *Statistics in Britain, 1865-1930: The social construction of scientific knowledge*. Edinburgh: Edinburgh University Press.
Nesselroade, J. R., Stigler, S. M., & Baltes, P. B. (1980). Regression toward the mean and the study of change. *Psychological Bulletin, 88*(3), 622-637.
Olby, R. (1989). The dimensions of scientific controversy: The biometric-mendelian debate. *The British Journal for the History of Science, 22*(3), 299-320.
Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. New York: Basic Books.
Pearson, K. (1902). On the fundamental conceptions of biology. *Biometrika, 1*(3), 320-344.
Pearson, K. (1930). *The life, letters and labours of Francis Galton, IIIA*Cambridge: Cambridge University Press.
Pence, C. H. (forthcoming). How not to fight about theory: The debate between biometry and mendelism in nature, 1890–1915. in The dynamics of science: Computational frontiers in history and philosophy of science, ed. G. Ramsey and A. De Block. Pittsburgh: University of Pittsburgh Press.
Pence, C. H. (2015). The early history of chance in evolution. *Studies In History and Philosophy of Science Part A, 50*, 48-58.
Porter, T. M. (1986). *The rise of statistical thinking, 1820-1900*. Princeton: Princeton University Press.
Porter, T. M. (2018). *Genetics in the madhouse: The unknown history of human heredity*. Princeton: Princeton University Press.
Provine, W. B. (1971). *The origins of theoretical population genetics*. Chicago: University of Chicago Press.
Radick, G. ((in press)). Disputed inheritance: The battle over mendel and the future of biology. Chicago: University of Chicago Press.
Radick, G. (2011). Physics in the galtonian sciences of heredity. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 42*(2), 129-138.
Shan, Y. (2020). *Doing integrated history and philosophy of science: A case study of the origin of genetics*. In Springer.
Stigler, S. M. (1986). *The history of statistics: The measurement of uncertainty before 1900*. Cambridge: Harvard University Press.
Stigler, S. M. (1999). *Statistics on the table: The history of statistical concepts and methods*. Cambridge: Harvard University Press.
Stigler, S. M. (2010). Darwin, Galton and the statistical enlightenment. *Journal of the Royal Statistical Society: Series A, 173*(3), 469-482.
Stigler, S. M. (2016). *The seven pillars of statistical wisdom*. Cambridge: Harvard University Press.

Stocking, G. (1987). *Victorian anthropology*. New York: The Free Press.
Waller, J. C. (2004). Becoming a darwinian: The micro-politics of sir Francis Galton's scientific career 1859–65. *Annals of Science, 61*(2), 141-163.
Weldon, W. F. R. (1902). Professor de Vries on the origin of species. *Biometrika, 1*(3), 365-374.
Weldon, W. F. R. (1906). Inheritance in animals and plants. In T. B. Strong (Ed.), *Lectures on the method of science* (pp. 81-109). Oxford: Clarendon Press.
Witteveen, J. (2019). Regression explanation and statistical autonomy. *Biology and Philosophy, 34*(5), 51.