

Karl Pearson and the Origins of Modern Statistics: An Elastician becomes a Statistician

M. Eileen Magnello

Introduction

A Renaissance scientist in Victorian London, Karl Pearson (1857-1936), was a prodigious and consummate literary polymath whose quest for philosophical, spiritual and numerical truth was his lifelong odyssey.¹ As a student of the Cambridge Mathematics Tripos system, Pearson learned to use applied mathematics as a pedagogical tool for determining the truth (i.e., 'one that provided the standards and the means of producing reliable knowledge'²). After graduating from Cambridge, Pearson spent a year studying in Heidelberg and Berlin in pursuit of the truth, which led him to write a passion play, some poetry, a romantic novel and a German book on aspects of the history of art. His passionate Germanic interests, which underscored his desire to find the truth, were pursued whilst he was writing papers and books on elasticity, engineering, mechanics, philosophy and physics, but the truth eluded him in these scientific fields: mathematical statistics was to become another means to that truth.

Pearson's legacy of establishing the foundations to contemporary mathematical statistics helped to create the modern world view, for his statistical methodology not only transformed our vision of nature, but it also gave scientists a set of quantitative tools to conduct research, accompanied with a universal scientific language that standardised scientific writing in the twentieth century. Despite his enduring contribution to statistics, Pearson had no preconceived ideas of becoming a statistician when he graduated from Cambridge in 1879. By the time he began to look for work in the 1880s, there were several options available to him: he could have become a mathematical physicist, an engineer, a lawyer, a specialist of medieval German literature, a philosopher or a mathematics teacher. The idea of creating a new mathematical discipline was not even on the horizon. Whilst I have been challenging existing accounts of Pearson for the last ten years and, in particular, redressing Weldon's influential role in Pearson's life,³ this paper will examine those factors that led Pearson to become a statistician and, in the process, to go on to create the foundations of modern mathematical statistics, having started his career initially as an elastician (that is, someone who devised mathematical equations for elastic properties of matter).

Historiography of Pearsonian Statistics

Much of the scholarship on Pearson has been shaped by mono-causal and uni-dimensional accounts; thus, less effort has been made to produce a more nuanced and more balanced account of Pearson's professional life. Pearson has long been erroneously viewed as a disciple of Francis Galton who, it is believed, followed in his footsteps and merely expanded what Galton started. Consequently, it has been falsely assumed by

various scholars that Pearson's motive to create a new statistical methodology arose from problems of eugenics.⁴ As I have already argued, Pearson not only managed the Drapers' Biometric and the Galton Eugenics laboratories separately, which occupied separate physical spaces, but he maintained separate financial accounts, established very different journals and created two completely different methodologies. Moreover, he took on his work in the Eugenics Laboratory reluctantly and primarily as a personal favour to Galton. Pearson thus emphasised to Galton that the sort of sociological problems that he was interested in pursuing for his eugenics programme were markedly different from the research that was conducted in the Biometric Laboratory.⁵ Conversely, many statisticians who have written on Pearson have invariably assumed that the impetus to Pearson's statistics came from his reading of Galton's *Natural Inheritance*.⁶ This view, however, fails to take into account that Pearson's initial reaction to Galton's book in March 1889 was actually quite cautious. It was not until 1934, almost half a century later, when Pearson was 78 years old, that he reinterpreted the impact Galton's book had on his statistical work in a more favourable light — long after Pearson had established the foundations to modern statistics.

Additionally, nearly all historians of science have failed to take into account that Pearson and Galton's ideas, methods and outlook on statistics were as different from each other as two people could have possibly been. Whilst Pearson's main focus was goodness of fit testing, Galton's emphasis was correlation (though Galton never even used Pearson's product-moment correlation); Pearson made higher level mathematics (i.e., determinantal algebra) a requisite for doing statistics and his work was thus more mathematically complex than Galton's; Pearson was interested in very large data sets (more than 1000) whereas Galton was more concerned with smaller data sets of around 100 (owing to the explanatory power of percentages) and Pearson undertook long term projects over several years, whilst Galton wanted faster results. Moreover, Galton thought all data had to conform to the normal distribution, whereas Pearson emphasised that empirical distributions could take on any number of shapes.

The idea that all data had to conform to the normal distribution was rooted in the philosophical ideologies of determinism and Aristotelian essentialism. Galton's belief in essentialism, which was the dominant thinking of the taxonomists, typologists and morphologists until the end of the nineteenth century and gave rise to the morphological concept of species, implied that species regressed back to the mean value. Galton was, therefore, convinced that all biological data could only be normally distributed. He was, in fact, so committed to his idea of a ubiquitous normal curve, that he created a mechanical device, a modified pantograph, to stretch or squeeze any figure in two directions until it was normally distributed. The Belgian statistician, Adolphe Quetelet (1796-1874) attached so much significance to the normal curve because of his belief in determinism, which meant that there was an ideal statistical mean value and that the normal curve was the ideal curve, since it followed the law of errors; hence, all the variation around the mean had to conform to this curve. That Pearson would go on to create a new type of statistics was to a large extent in response to the unshakable conviction, held by so many vital statisticians, mathematicians and philosophers, that the normal distribution was the only feasible distribution for the analysis and interpretation of statistical data. Such was the tyranny of the normal curve, that by the end of the nineteenth century, most statisticians assumed that no other curve could be used to describe data, but this monolithic view was challenged by Pearson in the last decade of the nineteenth century.

There has also been a tendency to overemphasise the role that Pearson's iconoclastic and positivistic book, the *Grammar of Science*, played in the development of Pearson's statistical work, whilst neglecting other influential factors in his life.⁷ This book, which contains Pearson's first eight Gresham lectures, written when he was 34 years old represents his philosophy of science as a young adult, and does not reveal everything about Pearson's thinking and ideas, especially those in connection with his development of the

modern theory of mathematical statistics. Thus, it is not helpful to see this particular book as an account of what Pearson was to do throughout the remaining 42 years of his working life. This uni-dimensional account of Pearson's work belies the complexity of Pearson, who was a far more multifaceted person than has been conveyed by a number of historians and philosophers of science.

W.F.R. Weldon and Darwinian Variation

The person who was, undoubtedly, the most influential individual in Pearson's life was the Cambridge-trained Darwinian zoologist, Walter Frank Raphael Weldon (1860-1906), known as 'Raphael' to Pearson, who provided Pearson with the much-needed impetus, data and indefatigable support that enabled Pearson to change careers from being an elastician to becoming a biometrician and, consequently, to construct a new statistical methodology.⁸ Whilst Galton did, indeed, play a role in the development of some of Pearson's statistical methods, mainly for correlation and regression, his role was not as significant as that of Weldon. Moreover, it was Weldon who introduced Pearson to Galton in 1894, three years after Pearson and Weldon first met. I will argue that the reasons Pearson was able to help Weldon in 1893, which led to Pearson's creation of the new discipline of mathematical statistics, were due to the following five factors: (1) Pearson's training for the Cambridge Mathematics Tripos examination, where he learned to use mathematics as a pedagogical tool for finding the truth, which set him on a life-mission to find this truth; (2) his disillusionment with physics in the late 1880s; (3) his belief that his earliest statistical ideas and methods that he devised in his Gresham Lectures from 1891 to 1894 could be applied to problems of evolution in 1892; (4) the fortuitous timing of Weldon's statistical questions to Pearson in 1892 – just when Pearson was beginning to devise new statistical methods and (5) Pearson's quest to achieve immortality by leaving a legacy that would survive him.

Weldon occupied a central position in Pearson's life, which can be easily documented from one of the most extensive sets of letters in Pearson's archives, which consist of nearly 1,000 letters of Pearson, Weldon and his wife Florence. Pearson and Weldon's letters tell the story of an intellectual love affair that was in full bloom throughout the 1890s, when Charles Darwin's ideas had kindled their intellectual partnership, when they began to look for empirical evidence of natural selection and a way to understand the circumstances needed for the formation of new species (or speciation). Once Darwin introduced the idea of continuous variation into biological discourse, his cousin, Francis Galton, began to devise statistical methods to measure this variation. Galton's work captured the interest of Weldon.

Pearson recognized the fundamental statistical concepts in Darwin's work, as 'every idea of Darwin, from variation, natural selection, inheritance to reversion, seemed to demand statistical analyses'.⁹ Darwin had not only shown that variation was measurable and meaningful by emphasising statistical populations rather than focusing on one type or essence, but he also discussed various types of correlation that could be used to explain natural selection. As the evolutionary biologist, Sewall Wright (1899-1988) remarked in 1931, 'Darwin was the first person to effectively view evolution as primarily a statistical process in which random heredity variation merely furnished the raw material.'¹⁰ Pearsonian mathematical statistics was thus built upon Charles Darwin's recognition that species comprised different sets of 'statistical populations' underpinned by individual *variation*. Biological Darwinism not only rejected the essentialistic concept of species, but it precipitated a new way of thinking that led Pearson and Weldon to create a new methodology.

Weldon first met Pearson in the autumn of 1891 at the Association for Promoting a Professorial University for London. From 1891 until Weldon took up his post at Oxford in 1899, they saw each other almost daily and often several times a day. Pearson not only regarded Weldon as 'one of the closest friends he ever

had' and valued his opinions more highly than those of anyone else, but by the time Pearson met Weldon, he found someone whose ideas meshed with those he had been developing in his Gresham lectures. Because Weldon's impact on Pearson led him to change the direction of his career, Pearson acknowledged Weldon as the person 'who changed the whole drift of my work and left a far deeper impression on my life' than anyone else.¹¹ Tragically, this intellectual love affair was cut short by Weldon's death in 1906 and left Pearson inconsolably bereft for the remaining 30 years of his life. To gain an understanding of the multifaceted development of Pearson, his early life and education, including his time in Cambridge and Germany, are addressed in the following sections.

Pearson's Early Life and Education

Carl was the younger son and second of three children; born in London, he was of Yorkshire descent, as most of his ancestors came from the North Riding. (The University of Heidelberg changed the spelling of his name in 1879 when they enrolled him as 'Karl Pearson'; he used both variants of his name interchangeably until 1884 when he finally adopted Karl, eventually becoming known as 'KP'.) His mother, Fanny Smith, came from a family of master mariners who sailed their own ships from Hull; his father William, who read law at Edinburgh, was a successful barrister and a Queen's Counsel (QC) at the Inner Temple of the Royal Courts of Justice. They were a family of dissenters and of Quaker stock; Carl's maternal grandfather was a Unitarian minister. By the time Carl was 22 he had rejected Christianity and adopted 'Freethought' as a nonreligious faith that was grounded in science, though he distinguished his views from a 'Freethinker' (i.e., a person who forms opinions about religion on the basis of reason without recourse to authority or established beliefs).¹²

His father was a very hard-working and taciturn man who was never home before seven in the evening. Once he was home he worked till midnight and was up at four in the morning reviewing his briefs. The only time he spent any time with his children was during the holidays. But to be home with their father was 'simply purgatory' because he never spoke a word to anyone. As a child, Carl was rather frail, often ill and prone to depression. Both Carl and his elder brother, Arthur found their father's attitude to be oppressive and they worried continually about their mother's well being.

When Carl was four, he had French lessons at their home on the Camden Road, opposite Holloway Gaol. After the family moved to Northwick, near Harrow-on-the-Hill in 1863 he and his brother, Arthur, attended a small school with 15 pupils in Harrow, established by a William Penn, who also provided home tuition for Carl in 1866. When the family moved to Mecklenburg Square, Bloomsbury later that year, the boys settled very happily into University College London School, then on Gower Street. Carl stayed for seven years, until he was 16 years old.

From the beginning, both parents wanted their sons to attend Cambridge, and at least one of them was expected to read mathematics. Since Arthur had received the Marlowe scholarship to read Classics at Trinity Hall, Cambridge, it was left to Carl to study mathematics. When Carl was 15 years old, his father began to look for a good Cambridge Wrangler to prepare him for the Mathematics Tripos examination. William's search was motivated because he was expecting far more from Carl than he was from Arthur, for he had thought that Carl was destined for a distinguished university career.¹³ At 16, Carl went up to Hitchin, 25 miles south-west of Cambridge, where he stayed for five months receiving tuition in mathematics. Very unhappy there, he left in July 1874 to go to Merton Hall, Cambridge to be coached in mathematics under a number of tutors including the great mathematics tutor, John Edward Routh, whom his father recommended to his son, since Routh 'had coached more Senior Wranglers than any other man'.¹⁴ Routh introduced Carl to the mathematical theory of elasticity, and started his hour-long tutorial

sessions with him, at seven in the morning. Carl stayed at Merton Hall from mid July 1874 until 15 April 1875 when he took his entrance examination at various Cambridge colleges. His first choice was Trinity College, where he failed the entrance exam choice; his second choice was King's College from whom he received an Open Fellowship. He began to study at King's on 9 October.

Being away from a depressing family life, Carl found the highly competitive and intellectually demanding requirements of the Mathematics Tripos to be the tonic he needed. He came to life in this environment and his health improved. In addition to studying, students of the Maths Tripos were expected to take regular exercise as a means of preserving a robust constitution and regulating the working day.¹⁵ Pearson carefully balanced hard mathematical study against such physical activities as walking, ice-skating, ice hockey and lawn tennis. When the weather was vile, for exercise he worked out indoors using dumbbells and played billiards. Success in competitive sport became a hallmark of the rational body whilst the hard study of mathematics was a manly pursuit.

As a diversion from mathematics, in Pearson's second year, he began to attend lectures on Dante and Spinoza. His mathematics tutor at Kings, Oscar Browning (1837-1923) recommended that Pearson read such Romantic works as Goethe's *Faust* and *Wilhelm Meister* and Percy Shelley's *Peter Bell the Third*. Inspired by Goethe's *The Sorrows of Young Werther*, perhaps the first European cult novel, Pearson 'was determined to write a book in the genuine gush style' and wrote his first book, a romantic novel, *The New Werther*.¹⁶ This was a literary work on idealism and materialism, written in the form of letters to his fiancée, Ethel, from a young man, Arthur, wandering in Germany and who, like Pearson, was searching for a creed of life and turned to philosophy, religion and history in the hopes of finding some underlying principle to life. Pearson graduated with honours in 1879, being the Third Wrangler in the Mathematics Tripos. He subsequently received a Fellowship from King's, which gave him financial independence for seven years. (He was made an Honorary Fellow of King's in 1903.) For the Victorians, being placed a high wrangler was a mark of enormous intellectual and social distinction.¹⁷ This was, in fact, such a high distinction that the names and the photographs of the top three Wranglers were published in all the national newspapers.

Pearson's German Wanderlust

Immediately after finishing the Tripos, Pearson began to plan his trip to Germany whilst pursuing other activities that interested him. Having already considered the possibility of becoming an elastician, Pearson began to spend three hours a day in Professor James Stuart's engineering workshop, whilst studying medieval languages, and reading philosophy – with the hopes of becoming a philosopher. He left for the continent in April 1879 to improve his German and to study physics and metaphysics. When Pearson was in Heidelberg, he read the works of Berkeley, Locke, Kant and Spinoza, but found his 'faith in reason has been shattered by the merely negative results which he found in these great philosophers that he despaired his little reason leading him to anything'.¹⁸ He subsequently abandoned philosophy because 'it made him miserable and would have led him to inevitably short-cut his career'.¹⁹ Later that summer Pearson was at such a 'low ebb of despair', in his search for the truth, that he was tempted to become a Roman Catholic. Whilst he wanted to know 'what is the truth?' he also realised that 'what is the truth for one, may not be the truth to another' man.²⁰ Angst-ridden, he wanted someone to tell him what his duty was, since he felt he had none.

This quest for the truth, which was of paramount importance to a Cambridge trained mathematician, became a template for the work he pursued and ultimately shaped the direction of his career. His time in Germany became a period of self-discovery: the romanticist and the idealist discovered positivism. Pearson

thus adopted and coalesced two different philosophical traditions to fulfil two different needs, for idealism was concerned with nature and personal feelings, whereas positivism dealt with science and professional goals.

Having relinquished philosophy, Pearson soon realised he would never be a great mathematician or physicist, like James Clerk Maxwell, William Thomson or Hermann von Helmholtz, because, as he put it, he 'was not a born genius'. His very good friend from King's College, Robert Parker, did not, however, accept Pearson's self-appraisal. Parker did not believe in born geniuses and admonished Pearson because 'for someone who had studied mathematics and philosophy you would have had as much a chance as most men of turning into a scientific God to be stuck on the mantelpiece of future generations of seedy undergraduates'.²¹ With some reluctance, Pearson decided to study Roman international law in Berlin in June 1879, even though he was still scribbling problems of mathematical physics. In fact, he was 'more or less distracted by mathematics during the whole of his time' in Heidelberg and Berlin.²² Whilst he had initially considered doing the Law Tripos at Cambridge, when he returned to London in June 1880 he decided instead to read law at Lincoln's Inn and took up rooms with Parker. Though Pearson had made this decision, admittedly under duress from his father, he was 'quite determined not to even open up a law book and still less to enter the law circuit' as his father proposed he should do, for he intended to 'plod on at old German' instead.²³

By October 1881, Pearson's father was becoming deeply worried about the unsettled life of his son, who occasionally needed money from him. William castigated Carl that it was 'high time you did something [with your life] as three years have already gone' and that 'you must take to reading law and only reading it'.²⁴ Pearson, however, wanted *another* six years before he did anything at the bar. Nevertheless, a month later he was called to the Bar, but after setting up a partnership deed between two turnip-top sellers in Covent Garden Market, which took him three agonising days to complete, he realised he 'hated the law'.²⁵ His mother was not surprised when Carl gave up the law, as she did not think that it was 'quite what his taste and inclination would really dictate'.²⁶ Looking for other opportunities, Pearson thought 'success might be a possible option on the road to science'.²⁷ With such a constant stream of activities and moving from one topic to another, Pearson confessed to his brother, Arthur, that 'I hardly know what to expect from myself, as I have so many different impulses which lead me in such opposite directions'.²⁸

Pearson's interests soon turned to German folklore and literature, the history of the Reformation and the German humanists. From 1882 to 1884 he lectured to working men's clubs around London. He lectured on heat in Barnes, on Martin Luther in Hampstead and on Karl Marx and Ferdinand Lassalle at Revolutionary Clubs around Soho. His lectures were, indeed, very popular and were often sold out. Many of his lectures on humanism and intellectual welfare in Germany were about his search for truth amongst religious systems, and he surmised that 'all systems of religion are of necessity half-truths'.²⁹ As a philological scholar of medieval German folklore, literature and its language, he was short-listed for the newly created readership in German, at Cambridge, in April 1884, except he 'longed to be working with symbols instead of words'.³⁰ By then he had already established himself as a respectable elastician and he began to write some papers on the theory of elastic solids and fluids, as well as some mathematical physics papers on optics and ether squirts.³¹

Mathematical Physics and the Quest for the Truth

As an elastician, some of Pearson's early work on elasticity involved, for example, determining the bending moments of a bridge span and calculating stresses on masonry dams. Since elasticity dealt with practical

problems and determining the geometry of space by examining surfaces within the body, Pearson's vision of the truth would have eluded him. Since an elastician cannot really see what is going on, except in a limited manner (e.g., at the point of rupture), any notion of the 'truth' would not have been physically obvious.

Ether, as Pearson explained, had been conceived by physicists to have been 'the medium that could fill up the interstices between bodies and between the atoms of bodies' in outer space.³² Pearson's theory of ether squirts was the final product from his theory of electromagnetism and atomism that he had been working on in the late 1880s. To Pearson, matter was geometry in motion, for it represented the changing shape of space. Pearson's ideas on ether squirts were influenced by the work of the philosopher and mathematician, William Kingdon Clifford (1845-1879), who presented the idea that matter and energy are simply different types of curvature of space and whose book, the *Common Sense of the Exact Sciences*, Pearson was asked to finish after Clifford's early death. Pearson developed his theory by combining two strands of ideas, one on pulsating spheres of ether and the other on Clifford's twists (or space curvatures) which Pearson regarded as comparable to magnetic induction. However, by the late 1880s, Pearson had become disillusioned with his idea of ether squirts, due largely to the lack of support he received from British physicists. His paper was eventually published in the *American Journal of Mathematics*.³³

Pearson thus concluded that it was not possible for the human mind to have knowledge of ultimate reality. The initial excitement that ether squirts could have provided him with the means to find that Victorian Cambridge idea of the truth, diminished when he realised he could not measure or weigh the ether squirts in outer space. Pearson abandoned this theoretical work and his interests returned to elasticity, if only to finish edit Isaac Todhunter's *History of the Theory of Elasticity*

Applied Mathematics and University College London

Between 1879 and 1884 Pearson applied for more than six mathematical posts in Dundee, Leeds, London, Manchester and Sheffield. Having no luck in finding a job that satisfied him, he thought of taking up a secretaryship in a hospital, becoming a school master, immigrating to North America or even returning to the law. Pearson took on a temporary job teaching mathematics at King's College, London in 1883. He accepted the Goldsmid Chair of 'Mechanism and Applied Mathematics' in October 1884, at University College London succeeding the German mathematician, Olaus Henrici (1840-1918). Pearson played a pivotal role in the institutional changes at University College London, as he created the Department of Structural (now Civil) Engineering in 1892, established a Department of Astronomy in 1904 with two observatories (the Transit and the Equatorial Houses) and founded the Biometric School in 1893, which was incorporated into the Drapers' Biometric Laboratory in 1903 and became the Department of Applied Statistics in 1911 (now the Department of Statistical Sciences). He also helped to establish the departments of anthropology and genetics. Over a period of 28 years, he founded and edited seven academic journals of which *Biometrika* is the best known periodical.

Though Pearson had finally landed a permanent job, he was unhappy a month after he started to teach. He lamented to Robert Parker that 'if I only had a spark of originality or was a genius, I would have *never* have settled down to the life of a teacher, but instead would have wandered through life in the hope of producing something that might survive me'.³⁴ But Parker could see all that Pearson wanted out of life was to be comfortable, having a little work and plenty of cash to make him financially independent.

Nevertheless, Pearson captivated the interest of his many engineering students; he lectured to groups, ranging from 80 to 100 students, for 11 hours a week. He felt there was a sense of power and inspiration in maintaining order among so many high-spirited young men, especially when failure would have meant

riots. His colleague, the eminent chemist, Sir William Ramsay (1852-1916), once remarked to Pearson that they 'were the only men who [could] hold big classes in complete silence in the College' .³⁵

Pearson continued to inspire his students in the twentieth century after he had established the discipline of mathematical statistics. According to a number of these students, they thought Pearson had the rare gift of complete clarity, coupled with an understanding and appreciation of what the student was going through. Moreover, Pearson showed a willingness to take the time to explain an idea so completely by numerical example that anyone could understand the lessons so long as they were willing to do some hard thinking.

The Gresham Lectures of Statistics

Pearson was, however, an ambitious man: shortly after his marriage in 1890 he took up the Gresham Chair of Geometry at Gresham College in the City of London. When Sir Thomas Gresham (1519-1579) founded his College, he established seven professorships on the lines of an old mediaeval university in which all knowledge fell into one of the seven divisions: divinity, astronomy, geometry, music, law, physic and rhetoric. The early occupants of the Gresham Chairs in Geometry and Astronomy, such as Christopher Wren, Robert Hooke, Robert Boyle and William Petty, were among the most distinguished scientific men of their time.

Pearson delivered a total of 38 lectures beginning in the spring of 1891 and ending in the summer of 1894; there were also five guest speakers (including Weldon) who delivered lectures when Pearson was ill. He eventually had to give up this post because his doctor recommended that he cut back on the amount of work he was doing. Pearson held this post for three years, concurrently with his post at UCL. These lectures were aimed at a very different audience from his engineering students at UCL. With the increasing development of the specialisation of academic disciplines in Victorian universities, there was a concomitant development of vocational education, which provided practical education and training to the working class through such organisations as the mechanics institutes, mining schools and agricultural programmes. Thus, students who attended Pearson's Gresham lectures consisted of the industrial class, artisans, clerks and others who worked during the day in the financial district in the City of London.

Pearson wanted to introduce these students to a way of thinking that would influence how they made sense of the physical world. Whilst his first eight lectures formed the bases of the *Grammar of Science*, the remaining 30 lectures dealt with statistics, which represent Pearson's earliest ideas about mathematical statistics. As he could not lead these students through the 'mazy paths of mathematical theory', he had to find a way to present material that would be accessible to this audience. He chose statistics as a topic for these students for he thought they would understand insurance, commerce and trade statistics and could relate to such games of chance involving Monte Carlo Roulette, lotteries, dice and coins. He appealed to them by using graphs, geometric figures and illustrations to teach statistics and deduction by easy arithmetic.

Pearson explained to these students, that 'the geometry of statistics was not just about the graphical representation of data, but it also was a fundamental process of statistical inquiry'. For Pearson, 'geometry was a mode of ascertaining the numerical truth and a means of statistical research'.³⁶ Furthermore, he proceeded to redefine statistics to these students because statistics did not have to be confined to vital statistics or the measurement of social phenomena, as Pearson argued that statistics was not a branch of sociology, but rather an abstract science in its own right. Thus, it seemed to him that

it is clear that a great future awaits our present statistics and we may reasonably anticipate that the combination of statistics and analysis will create a science which will excel every other branch of

mathematics, including astronomy, mechanics and physics.³⁷

He recognised that statistics could be useful for problems in biology, to measure variation, Darwinian natural selection and for problems of heredity.

Thus, we can learn far more about Pearson by looking at *all* of his Gresham lectures, rather than simply looking only at the first eight published lectures, which formed the basis of the *Grammar of Science*. His 30 statistical Gresham lectures not only provided Pearson with the opportunity to create a new statistical methodology, but they signified a turning point in Pearson's career, for he was able to pursue his goal of finding the numerical truth when he began to teach the geometry of statistics.

Though Pearson was to find himself on the threshold of creating a new kind of statistics at the end of 1891, it was not until Weldon asked him for his advice on the data from his Naples crabs in 1892 that Pearson's early statistical ideas came to fruition. When Weldon was a student at UCL, before going up to Cambridge to read zoology, he acquired a respectable knowledge of mathematics from Pearson's predecessor, Olaus Henrici, whose emphasis on the use of graphical methods to teach mathematics, may have shaped Weldon's use of graphical procedures when analysing some of his statistical data in the early 1890s. This also may have helped foster his symbiotic relationship with Pearson who used geometry as a heuristic device to teach statistics. From 1892 to 1897, Weldon and his wife Florence travelled during the summer holidays to Guernsey, Rome, the south of France and the Bahamas to collect marine biological data. Before moving to London from Cambridge, in the autumn of 1889 to take up his new post in the Jodrell Chair of Zoology at UCL, Weldon went to Plymouth that summer to collect data on marine organisms and he began to use Galton's statistical methods after he read *Natural Inheritance*. Weldon had first met Galton in Swansea some nine years earlier at the annual meeting of British Association for the Advancement of Science.

Pearson had just devised the standard deviation in 1892 when Weldon approached him for assistance because he found that one of his distributions of data was bimodal, while the rest of his data were normally distributed. Weldon's attempt to break up his double humped curve into two normal components was derived from Galton's belief that all distributions should be normally distributed. Weldon then concluded that either the crabs from Naples were two distinct races or they were in the process of creating a new species. He also seems to have been exploring Galton's claim that a new species could be established only by a sport, jump or saltation producing a new type (i.e., instantaneous speciation). Pearson wanted to find another way to interpret the data without trying to normalise it, as Galton and Quetelet had done. Pearson and Weldon thought it was important to make sense of the shape of the curve without distorting its original shape, as it might have revealed something about the creation of new species. Thus, for him the truth could not be found by forcing data to conform to the normal curve.

Weldon wanted to measure the variation in an attempt to gain some understanding of evolution, to determine how a new species emerged and then to detect empirical evidence of natural selection. He needed a statistical system that could measure the variation of his crabs and one that could systematically handle large amounts of data, since he had collected thousands of crab measurements. Pearson realised that a very large sample is essential when trying to show empirical evidence of natural selection. Since Galton's samples were usually not larger than 100, his statistical methods were not amenable to Weldon's data. To help Weldon, Pearson had to create a formalised system of frequency distributions that could handle large sample sizes (e.g., more than 1,000 female crabs) and to develop a system that did not rely on the normal distribution.

After Pearson helped Weldon, it then became possible to make comparisons or generalisations with other data sets that had been previously impossible to make. Pearson had already introduced the 'histogram' on 18 November 1891, a term he coined to designate a 'time-diagram' in his lecture on 'Maps and Chartograms'. He explained that the *histogram* could be used for *historical* purposes to create blocks of time of 'charts about reigns or sovereigns or periods of different prime ministers'.³⁸ The histogram is similar to a bar chart, except the histogram is contiguous; it has no gaps, and is more faithful to exact data, whereas a bar chart has gaps and uses nominal data. By November 1893, Pearson realised that scientists were approaching the question of heredity and evolution from a new standpoint. Thus, he pronounced that 'for the first time in biology, there is a chance of the science of life being an exact, a mathematical science' and that it was 'largely to Weldon that we owe this attempt to give an exact aspect to the problem of evolution: his intensely laborious and careful measurement on the organs of shrimp and crab are the first step in the right direction'.³⁹

Goodness of Fit Testing

Once Pearson began to help Weldon with his data, he began to import Cambridge maths into his statistical methods and theory. He adapted the mathematics of mechanics, using the *method of moments*, to construct a new statistical system to interpret Weldon's data that produced asymmetrical curves, since no such system existed at the time. The term 'moment' originated from mechanics and is a force about a point of rotation, whilst moments in statistics are averages. From the method of moments, Pearson established four parameters for curve fitting to show how the data clustered (the mean), how it spread (the standard deviation), if there were a loss of symmetry (skewness) and if the shape of the distribution were peaked or flat (kurtosis). These four parameters describe the essential characteristics of any distribution: the system is parsimonious and elegant. The method of moments, which enabled Pearson to create the infrastructure of his statistical methodology, are *still* essential for interpreting any set of statistical data, whatever shape the distribution takes. Hence, this system allowed Pearson to analyse data that resulted in various shaped distributions, and enabled him to move beyond the limitations of the normal curve.

After Pearson examined Weldon's asymmetric curves that had been derived from his crab data in Naples, Pearson decided that an objective method of measuring the goodness of fit was a desideratum for distributions that did not conform to the normal curve. Pearson's earliest consideration of determining a measure of the goodness of fit test came out of his lecture on 21 November 1893, when he asked his students, 'Can you always fit a normal curve to a set of data?' The answer was 'not always' since there were many types of data that could produce asymmetric curves and thus would not fit into a normal curve.⁴⁰ Pearson went on to devise his first goodness of fit test for asymmetrical distributions, using the sixth moment from the method of moments in 1892.

Quetelet had, in fact, made one of the earliest attempts to fit a set of observational data to a normal curve in 1840, which Francis Galton began to use in 1863. Wilhelm Lexis devised the *Lexican Ratio L* as a goodness of fit test to determine if an empirical distribution conformed to the normal distribution, whilst Francis Ysidro Edgeworth provided a goodness of fit test in 1887 that was based on a normal approximation to the binomial distribution.⁴¹ Though many other nineteenth century scientists attempted to find a goodness of fit test, such as the American statistician, Erasmus Lyman de Forest, and the Italian mathematician, Luigi Perozzo, they did not give any underlying theoretical basis for their formulas, which Pearson managed to do

By the time Pearson had finished his 30 statistical Gresham lectures in May 1894, he had provided the underpinnings of his statistical methodology and he was in the process of creating a new academic

discipline. With such well-attended lectures, his public life was well developed by the time he began to teach statistics at UCL. He was thus able to bring to UCL what he gained at Gresham. In October, he began to offer a set of lectures at UCL on the 'theory and Practice of Statistics' for one hour a week for those 'desiring to study Animal Variation, to deal with Errors of Physical Observations or to become Actuaries'.⁴² These lectures were, for Pearson, 'not a part of his regular duty, but solely instituted because [he] was interested in developing a modern theory of statistics'.⁴³

In the last set of his Gresham lectures in May 1894, Pearson discussed various procedures for goodness of fit testing for asymmetrical curves. He introduced his second measure of a goodness of fit test at UCL in the spring of 1894, which he thought provided a 'fairly reasonable measure of a goodness of fit'. He continued to work on improving this method throughout the 1890s until he devised his chi-square goodness of fit test in 1900. However, before he reached this solution in 1900, his work was interrupted by Francis Galton who needed some help with correlation. The interruption proved beneficial, as Pearson was able to expand the corpus of his statistical methods; he went on to devise 22 methods of correlation of which 11 continue to be used today.

On Christmas Day in 1896, Pearson wrote to Galton that he wanted to develop a goodness of fit test for asymmetrical distributions for biologists and economists, which meant that Pearson also needed to create a corresponding probability distribution that was asymmetrical in shape. Pearson's ongoing work on curve fitting signified that he needed a criterion to determine how good the fit was, which led him to devise different goodness of fit tests. This work underpinned the infrastructure to his statistical theory and encompassed his entire working life as a statistician; this began in 1892 when he introduced the sixth moment as a measure of goodness of fit for Weldon's data, continued throughout the 1890s, culminated in 1900 when he found the exact chi-square distribution from the family of Gamma distributions and devised his *chi-square* (χ^2 , P) *goodness of fit test* and ended with the last paper he wrote when he was 79 years old. Indeed, the chi-square goodness of fit test represented Pearson's single most important contribution to the modern theory of statistics, for it raised substantially the practice of mathematical statistics. The overriding significance of the chi-square distribution and its goodness of fit test meant that statisticians could use statistical methods that did not depend on the normal distribution to interpret their findings.

Pearson also established the professional accoutrements necessary to establish and to institutionalise the new discipline of mathematical statistics: with Weldon he founded *Biometrika* in November 1900, he established the Drapers' Biometric Laboratory in 1903, and set up the first-ever degree course in mathematical statistics in 1917. Largely due to Weldon's assistance, and later that of his many students, including George Udny Yule and William Sealy Gosset (or Student) as well as Francis Galton, Pearson was able to firmly establish the new discipline of mathematical statistics in the early years of the twentieth century, which provided the foundations for such statisticians as Ronald A. Fisher to make further advancements in the development of the modern theory of mathematical statistics.

Notes

¹ This paper is based largely on material extracted from my forthcoming book, *Victorian Values: Karl Pearson and the Origins of Modern Statistics*.

² Andrew Warwick, (Personal Communication, 4 March 2004).

- 3 See Eileen Magnello, 'Karl Pearson's Gresham Lectures: W.F.R. Weldon, speciation and the origins of Pearsonian statistics', *British Journal for the History of Science*, 29 (1996), 43-63; *Idem.*, 'Karl Pearson's Mathematisation of Inheritance: From Ancestral Heredity to Mendelian Genetics (1885-1909)', *Annals of Science*, 55 (1998), 35-94.; *Idem.*, 'The Non-Correlation of Biometrics and Eugenics: Rival Forms of Laboratory Work in Karl Pearson's Career at University College London', *History of Science* 37 (1999) 79-106; 123-150 and *Idem.*, 'The Introduction of Mathematical Statistics into Medical Research: The roles of Karl Pearson, Major Greenwood and Austin Bradford Hill' in Eileen Magnello and Anne Hardy (eds.), *The Road to Medical Statistics*. (New York and Amsterdam: Rodopi, 2002) 95-124.
- 4 Daniel Kevles, *In the name of Eugenics* (New York, 1985), p. 39; Donald Mackenzie, *Statistics in Britain: The Social construction of Scientific Knowledge 1856-1930*. (Edinburgh 1980), p.180; Theodore Porter, *The Rise of Statistical Thinking* (Princeton, 1986) p. 305; Richard Solloway, *Demography and Degeneration* (Chapel Hill, 1990), 117-18.
- 5 Eileen Magnello, 'The Non-correlation of Eugenics and Biometrics' *History of Science* (1996).
- 6 For a recent account **see** Michael Bulmer. *Francis Galton:: Pioneer of Heredity and Biometry*, (Baltimore: Johns Hopkins University Press, 2003)
- 7 This view was espoused initially by Bernard Norton, 'Biology and Philosophy: The methodological foundation of Biometry' *Journal of the History of Biology*, 7 (1975) 85-93, which took on an even greater role in the work of Theodore Porter, especially in his recent book, *Karl Pearson: The Scientific Life in a Statistical Age*, (Princeton University Press, 2004).
- 8 I made this point in 'Karl Pearson's Gresham Lectures: W.F.R. Weldon, speciation and the origins of Pearsonian statistics', *British Journal for the History of Science*, 29 (1996), 43-63 and 'The Non-Correlation of Biometrics and Eugenics: Rival Forms of Laboratory Work in Karl Pearson's Career at University College London', *History of Science* 37 (1999) 79-106; 123-150.
- 9 Karl Pearson, 'Editorial. (II). The Spirit of *Biometrika*' , *Biometrika*, (1 October 1901) p. 4.
- 10 Sewall Wright, 'Evolution in Mendelian Populations' *Genetics*, 16 (1934) 97-159.
- 11 Karl Pearson, Letter to W.F.R. Weldon, 28 February 1899, KP:UCL/266/8.
- 12 See Theodore Porter, *Karl Pearson. The Scientific Life in a Statistical Age*. (Princeton University Press, 2004), p. 108.
- 13 William Pearson, Letter to Carl Pearson, 14 December 1873. Archival material assembled by Helga Hacker Pearson at University College London, /Tu.73.HV; Hereafter referred to as HHP:UCL.
- 14 William Pearson, Letter to Carl Pearson (30 April 1873). Archive of Karl Pearson at University College London; Hereafter referred to as KP:UCL
- 15 Andrew Warwick, 'Exercising the Student Body. Mathematics and Athleticism in Victorian Cambridge' in Steven Shapin and Christopher Laurence (eds.). *Science Incarnate: Historical Embodiments of Natural Knowledge*. (Chicago: University of Chicago Press, 1998).
- 16 For a detailed analysis of Pearson's *New Werther*, see Theodore Porter, *Karl Pearson: The Scientific Life in a Statistical Age*, (Princeton University Press, 2004).

- 17 Andrew Warwick, 'The World of Cambridge Physics' in Richard Staley (ed.) *The Physics of Empire: Public Lectures*. (Whipple Museum of the history of Science: Cambridge, 1994) p.61.
- 18 Carl Pearson , Letter to Robert Parker, 17 August 1879, HHP:UCL/K/F. H.24.
- 19 Carl Pearson, Letter to Robert Parker 19 October 1879. KP:UCL/922.
- 20 Carl Pearson, Letter to Robert Parker, 20 June 1879. KP:UCL/922.
- 21 Robert Parker Letter to Carl Pearson, 26 October 1879. HPH:UCL/K/R.79.B.10-12
- 22 Carl Pearson, Letter to Fanny Pearson, 1 March 1880 (HPH:UCL/K/F/80.B29-30.
- 23 Carl Pearson, Letter to Robert Parker, 1 October 1881, KP:UCL/922/2.
- 24 Pearson complained about his father's ranting in a letter to Robert Parker (24 October 1881) KP: UCL/922.
- 25 Carl Pearson, Letter to Robert Parker, 8 October 1881 KP:UCL/922.
- 26 Fanny Pearson, Letter to Carl Pearson, 7 November 1879. HHP:UCL/K/R.79.B.22
- 27 Carl Pearson, Letter to Robert Parker, 12 Jan 1880. KP:UCL/922.
- 28 Carl Pearson, Letter to Arthur Pearson, 18 May 1879. HHP: UCL/K/F.79.H.16-19
- 29 Karl Pearson, *The Ethic of Freethought and other Addresses and Essays* (London: T. Fisher Unwin, 1888).
- 30 Karl Pearson, Letter to Oscar Browning, 16 January 1880, Oscar Browning Collection, King's College, Cambridge University.
- 31 Carl Pearson, 'On the motion of spherical and ellipsoidal bodies in fluid media' *Quarterly Journal of Pure and Applied Mathematics*, 20: 60-80; Part II: 184-211; *Idem.*, 'Note on twists in an infinite elastic solid' *The Messenger of Mathematics*, 13, 79-95. *Idem.*, 'On a certain Atomic Hypothesis' *Transactions of the Cambridge Philosophical Society*, 14: 71-120 (Read 2 February 1885).
- 32 Karl Pearson, *The Grammar of Science* (Charles and Black: London, 3rd edn, 1911). 205.
- 33 Karl Pearson, 'Ether Squirts. Being an attempt to specialize the form of ether Motion which forms an Atom in a Theory propounded in former papers' *American Journal of Mathematics*, 13, (1891), 309-362.
- 34 Karl Pearson, Letter to Robert Parker, 12 November 1884, KP:UCL/922/2.
- 35 Egon Pearson, *Karl Pearson An Appreciation of some aspect of his life and work* (Cambridge University Press, 1938), p. 16. Pearson made this comment in a letter to his son, Egon.
- 36 Karl Pearson, Gresham Lecture , 'Geometry of Statistics' (17 November 1891), KP:UCL/49
- 37 Karl Pearson, Gresham Lecture , 'Geometry of Statistics' (17 November 1891), KP:UCL/49
- 38 Karl Pearson, Gresham Lecture on 'Maps and Chartograms' (20 November 1891) KP:UCL/49/21.
- 39 Karl Pearson, Gresham Lecture , 'Compound Curves' (24 November 1893),), KP:UCL/49
- 40 *Ibid.*

41 F.Y. Edgeworth, 'The empirical proof of the law of errors' *Philosophical Magazine* (Fifth Series) 24, 330-47.

42 Anon, 'The Retirement of Karl Pearson', *UCL Magazine* (Summer, 1933), 166.

43 Karl Pearson, Report to the Court of the Worshipful Company of Drapers' (1918) KP:UCL #233. p.4