Karl Pearson: perceptions, statistics and scientific objectivity

Francesco Masedu, Marco Valenti

Department of Computer Science, University of L'Aquila, Italy

Corresponding Author:
Francesco Masedu, Department of Computer Science, University of L'Aquila
Via Vetoio (Coppito) - 67100 - L'Aquila, Italy
E-mail: francesco.masedu@di.univaq.it

Summary
This article looks briefly at the life of the great British statistician Karl Pearson, and outlines his ideas on the use of probability and statistics to categorize sequences of perceptions. Pearson's aim was to give scientific knowledge a foundation, making it the only legitimate form of real knowledge.

KEY WORDS: *probability, scientific method, objectivity.*

Introduction

This discussion addresses the life and work of Karl Pearson, who must be recognized as one of the main founders of modern biostatistics; it examines his conception of scientific knowledge and the way judgements are based ultimately on sequences of perceptions and their uniform statistical character.
The above topic is dealt with mainly in *The Grammar of Science* (1), where we find Pearson's position expounded in detail. This book will be taken as a sort of main thread as we seek to clarify Pearson's epistemic assumptions, without claiming to be exhaustive, and his idea on the critical role played by probability in establishing genuine scientific propositions. Despite the time that has elapsed since its publication, *The Grammar of Science* still makes good reading for those interested in understanding the position of probability and statistics in modern science and "It is beyond question a remarkable book for boldness of thought and erudition. It is the strongest of advocates for scientific study, as possessing alone the power of making a man think without bias" (2).

Fundamental assumptions: the problem of data

A look at the biography (3) of Karl Pearson (London, 27th March 1857 – Coldharbour, Surrey, 27th April 1936) immediately reveals a man who had a complex and eclectic education, which began in Cambridge, where he studied mathematics and physics under teachers like Stokes, Cayley and Maxwell – Pearson graduated from Cambridge in 1879 as "Third Wrangler in Mathematical Tripos" – while also reading classics of philosophy and literature by authors such as Spinoza, Dante, Rousseau and Goethe. He was also interested in religious and philosophical subjects, which he discussed in depth with his fellow students (4). His scholarly interests, especially after leaving Cambridge for Germany – it was probably here that he began to use Karl instead of Carl – and before statistics became the main focus of his research, included German studies, Marxism and Roman law.
Crucial to his cultural development was the publication of *Natural Inheritance* (1889) by Francis Galton (5), which prompted a shift in his interests towards eugenics and statistics applied to inheritance (6). It is also worth mentioning his friendship with Walter

Frank Raphael Weldon (London, 15th March 1860 – London, 13th April 1906), a zoologist and statistician at University College London who helped to crystallize a number of the problems and issues that Pearson dealt with in the course of his career (7). At the age of 33, Pearson, a recognized scholar in many subjects, turned his attention specifically to statistics, as witnessed by the works of this period, namely, his series of papers entitled *Mathematical Contributions to the Theory of Evolution*. Although this is just a very brief summary of Pearson's scientific life we cannot fail to mention his foundation, with Galton and Weldon, of *Biometrika*, the journal he edited until his death in 1936; among his many achievements, Pearson must also be remembered for developing the notion of correlation, for characterizing the idea of fitting curves to observed data using the famous Chi-square (8), and for creating fundamental instruments such as the *Tables for Statisticians and Biometricians* (1914) (9). In accordance with the prevailing doctrine of his age – we refer to positivism, which claimed that the objectivity of scientific facts is the only real guarantee of knowledge and that only science can advance legitimate cognitive propositions –, Pearson turned to the application of quantification and quantifiability, particularly to biosciences, as the main focus of his activity.

He pursued four main, mutually dependent, goals:

a) to reduce knowledge to scientific knowledge, on the basis of its quantifiability ;
b) to express objectivity in mathematical-quantitative terms;
c) to eliminate the scientist from the data he deals with;
d) finally – this aim is related to the previous ones –, to arrive at the necessary classification of "facts", turning them into "scientific facts" susceptible to absolute judgements.

In Pearson's own words "The classification of facts and the formation of absolute judgements upon the basis of this classification – judgements independent of the idiosyncrasies of the individual mind – is peculiarly *the scope and the method of modern science*. The scientific man has above all things to aim at self-elimination in his judgements, to provide an argument which is as true for each individual mind as for his own. *The classification of facts, the recognition of their sequence and relative significance is the function of science*, and the habit of forming a judgement upon these facts unbiased by personal feeling is characteristic of what we shall term the scientific frame of mind" (1).

Starting from the premise that the constituents of knowledge are immediate sensory impressions and memory, which thus implicates both the physical and the mental facts that science deals with, the English statistician faced two key issues: how can the subjectivity of sensory impressions and the judgements forming scientific propositions reasonably be linked? And what is an "external object"?

We always find, at the root of any representation of the world, immediate sensory impressions that, through an process of associations, are appraised on the basis of their similarities with past impressions, as recorded in the memory.

The idea is that humans link immediate sensory impressions to a "subject", defining the reality of an object as the likelihood of its occurring in a group of immediate sensory impressions. The definition in this case is nominal, in the sense that it is not necessary to postulate the thing in itself; in Pearson's view, our assumption of the world outside amounts to a mere metaphysical illusion. It must be noted, however, that this "illusion of the world outside" emerges again, as so often happens in extreme forms of empiricism, in the attribution of science with an ontological value so that we might approach it, but not "beyond the sense impressions, beyond the brain terminals of the sensory nerves" (10). Pearson seems to combine a positivist setting with empiricism that, in order to justify the intersubjective universality of scientific statements, accommodates a moderate kantism so that: "The same type of physical organ receives the same sense-impressions and forms the same constructs. Two normal perceptive faculties construct practically the same universe" (1).

If we allow (without necessarily accepting) this questionable means of excluding the metaphysical "immaculate purity of the facts", objectivity is established on the basis of the principle of the uniformity of perceptual faculties, or in the words of the English scientist: "The universal validity of science depends upon the similarity of the perceptive faculties and reasoning in normal civilized men" (1).

This concept, with its constant reference to *normal civilized men*, leads to the solution of an issue every bit as big as the one that it sets out to resolve, appearing quite ideological as well as influenced by Pearson's studies on eugenics. These studies form an apparatus, mostly implicit, fundamental in determining the concept of "normalcy" that provides the criterion for recognizing, on a statistical basis, the objectivity of data.

The results of the Pearson's gnoseology emerge quite explicitly in his social Darwinism and idea of evolution as he theorizes: "History shows me one way, and one way only, in which a high state of civilization has been produced, namely, the struggle of race with race, and the survival of the physically and mentally fitter race" (11) and also "You cannot get a strong and effective nation if many of its stomachs are half fed and many of its brains untrained. We, as a nation, cannot survive in the struggle for existence if we allow class distinctions to permanently endow the brainless and to push them into posts of national responsibility" (12).

So, the facts of science become groups of stored sensory impressions or, at least in Pearson's desiderata, conceptual constructs that, carrying long chains of inferences, are no longer directly verifiable in an immediate perceptual reference or in memory.

Objects and classification of sensory impressions

But what must ("must" to indicate the regulatory spirit of Pearsonian epistemology) the concrete criteria of validity that a scientific concept should meet actually be; Pearson lists two:

The criterion of self-consistency: an idea is to be considered non-contradictory when it is connected to possible sensory impressions (e.g. the idea of winged horse, unlike the idea of square circle, is not contradictory). Here, we are on the same wavelength as Hume, who had distinguished between matters of fact and relations of ideas (13).

The criterion of intersubjectivity: the idea appeals to every "normal" individual's store of sensory impressions.

The debt owed to Hume is clear, even though the points of contact with Mach are more interesting. We refer to the concept that the scientific method, which stands out for its effectiveness and economic character, is the only possible source of knowledge. A scientific law "… will be accepted by every rational mind which has once understood its terms and clearly analysed the facts which it resumes" (1). A scientific law, then, is meaningful only in relation to human perceptual and rational faculties, it cannot and does not aim to represent an *ordo naturae*; it speaks out not on the *world*, so much as on that which, because of the faculties mentioned, *is present in men as the world*.

Rationality must classify sequences of sensory impressions, making comparisons in order to structure and to clarify that which takes place within us, and in this process mathematics certainly has an essential role, but a limited one: we could liken it to a sort of mental shorthand. For Pearson, a scientific law is a descriptive summary of perceptual sequences, a summary that does not express a simple concatenation of phenomena, but rather must correlate concepts constructed using generalizations, and the picture that emerges is one of empiricism giving way to a moderate kantism (14). In this sense, Lenin, too, in his *Materialism ad Empiriocriticism* correctly identified and outlined this aspect of Pearson's thought. According to Haldane, "Now Lenin disagreed strongly with Pearson, and claimed, in my opinion correctly, to have found self-contradictions in his arguments. Nevertheless, he found him vastly clearer than other Machians. Let me read a few of Lenin's sentences. 'The philosophy of Pearson, as we shall repeatedly find, excels that of Mach in integrity and consistency' (p. 119). 'The Englishman, Karl Pearson, expresses himself with characteristic precision, "Man is the creator of natural law" (p. 221). And finally (p. 243) Lenin described him as 'This conscientious and scrupulous foe of materialism'" (15).

According to the English statistician, science does not explain but describes, and necessity relates to concepts and not to perceptions, which brings us face to face, once again, with the human gnoseological nightmare about the validity and the role of knowledge.

Probability and the method of science

In the light of what we have said, one can hardly be surprised to discover that the style of research proposed by Pearson is Baconian, along the lines of the inductivist tradition that in England is associated with distinguished scholars like John Stuart Mill (16), so that, for him, research must always begin with the simple collection of facts.

There are two components that characterize Pearsonian inductivism: *uniformity of nature* guaranteed, as already seen, by the alleged *universal uniformity of our perceptual faculties* and also the *belief in the constant repetition of perceptual sequences,* which can now avail itself of the theory of chance. Let us consider two corollaries of this latter idea:

i) an established belief in the future uniform repetition of a perceptual sequence shifts the fundamental question to the epistemological status of probability and the annexed mathematical theory;
ii) given the non-necessity of sensory impressions, the very concept of causation dissolves into simple denotation (nominal character of the concept of cause) of connections between constant perceptual phenomena.

We have science only if there is perceptual uniformity, raised to the condition of knowability, or to be precise, of scientific knowability; in the case of *Erlebnisse*, of perceptual chaos, we simply do not have this condition.

One aspect that strongly characterizes the Pearsonian knowledge design is the *holism of inductive certainty*. This aspect is linked to the intriguing issue of the so-called rising sun problem.

Pierre Simon de Laplace (1749-1827) had raised an interesting issue: if we consider an event that has always taken place, in a sequence of n observations, what probability can be attributed to that same event occurring for the $n+1^{th}$ time? The great French mathematician, by making use of the principle of indifference and Bayes theorem, came up with the answer $(n+1)/(n+2)$ (17), thereby providing an estimate, somewhat controversial, of the probability that the sun will rise tomorrow (18). According to the English statistician, in the calculation of that probability we must take into account two factors:

1) n observations of the phenomenon;
2) the general repetition of perceptual sequences in all types of phenomena.

The holism we refer to here resides in the fact that Pearson criticizes Laplace's use of only the first factor (he ignores the second completely). Point 2) indicates that the predictive capacity of our knowledge, and any scientific knowledge must be predictive, depends upon the absolute totality of past experience. This is interesting because holism is a prominent issue, used in very sophisticated ways, in the thought of the twentieth century.

Pearson reformulates the problem as follows (1): let us suppose that we have experienced m perceptual sequences, each of which has been repeated n times without exception, and are given another sequence of s perceptions, repeated without exception, what is the likelihood that this latter sequence will be repeated for the $s+1^{th}$ time? From Laplace's formula and from point 2) Pearson gets $[m(n-1+s)]/[m(n-1)+s+1]$.

The essential point is that the perceptual sequences to which we are referring are those of the whole history of humanity, which makes the uniform repetition of the sequence a practical certainty.

Frankly, this point is disconcerting both on account of its impracticality and, above all, because of the absolute lack of clarity with regard to the meaning of perceptual sequence: what should we regard as the characteristics that make two perceptions distinct? Ultimately, Pearson's perceptual atomism is not clear at all, we do not have identity criteria for perceptions, and we could say, like W. O. Quine, "no entity without identity" (19).

Discussion

Pearson regarded the problem of the objectivity of perceptions, and its related statistical character, as the problem of the objectivity of scientific discourse. We have seen that for Pearson the "raw material of science" consists of stable sequences of perceptions, and that experimental observation and mathematical processing both aim to: establish "associations", developing adequate conceptual constructions regarded as "economic descriptions".

It is quite clear that, by stopping at the level of individual perceptions, no emphasis at all is placed on the constructive nature of scientific data. Like other forms of sensism, Pearson's design is confused and unable to cope with the transition from singular to universal propositions, so the exclusion of the subject from scientific knowledge is based on the fundamental and decisive role played by statistically weighted human impressions, maybe the most subjective aspect of a scientific fact. The difficulty mentioned emerges alongside other attempts (e.g. Ernst Mach, Bertrand Russell and Rudolf Carnap), interesting but unconvincing, to derive scientific facts from the perceptual flow.

The evidence of perceptual level, being confused, does not allow the identification of intersubjective degrees of stability for perceptual sequences, which would make it possible to move from *percepta* to concepts on which to operate mathematically.

We conclude our discussion about Pearson's conception of science and knowledge with the words of the great geneticist John Burdon Sanderson Haldane (1892-1964) who quotes sentences from the Grammar, which to go back to Pearson's own views, and in his view illustrate the strength and the weakness of Pearson's approach to science. "'The unity of all science consists alone in its method not in its material' 'No physicist ever saw or felt an individual atom. Atom and molecule are intellectual conceptions by aid of which physicists classify phenomena and formulated the relationship between their sequences'. The strength is shown by the fact that the distributions, which Pearson worked out to describe Weldon's measurements of populations of crabs, will serve equally well to describe populations of stars, manufactured goods, durations of life, incomes, barometer readings, and so on. The weakness is shown by the fact that physicists have, during this century, seen individual atoms, or rather atomic nuclei, by the tracks which they make when moving rapidly. Pearson's philosophy discouraged him from looking too far behind phenomena" (15).

References

1. Pearson K. The Grammar of Science. 1st ed. London: Walter Scott, 1892.
2. The New York Times, August 1892.
3. Edwards AWF. RA Fisher on Karl Pearson. Notes and Records of the Royal Society of London 1994; 48: 97-106.
4. Pearson ES. Karl Pearson: an appreciation of some aspects of his life and work: Biometrika 1936; 28: 193-257.
5. Galton F. Natural Inheritance. London and New York: MacMillan and Co., 1889
6. Stigler SM. Statistics on the Table: the History of Statistical Concepts and Methods, 2nd edition. Cambridge, Massachusetts: Harvard University Press, 2002:13.
7. Walker HM. Karl Pearson. International Encyclopaedia of the Social Sciences XI. New York: MacMillan, 1968: 496-503.
8. Pearson K. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Phil. Mag. 1900; 6: 157-175.
9. Pearson K. Tables for Statisticians and Biometricians. Cambridge: Cambridge University Press, 1914.
10. Nagel E. The Grammar of Science by Karl Pearson, Introduction. New York: Meridian Books, 1957: viii.
11. Pearson K. National Life from the Standpoint of Science, 2d Ed., Cambridge: Cambridge University Press, 1919: 21-22.
12. Karl Pearson, National Life from the Standpoint of Science (op.cit.): 43-47.
13. Hume D. An enquiry concerning human understanding. Originally published in London, 1777. Web edition: Leeds Electronic Text Centre, July 2000. Text prepared by Peter Millican from the edition held in the Brotherton Library, University of Leeds.
14. Sandrini MG. Proceedings of the Conference of History and Philosophy of Science. Lucca, 1993.
15. Haldane JBS. Karl Pearson, 1857-1957. Biometrika 1957; 44: 303-313.
16. Mill JS. A System of Logic, Ratiocinative and Inductive: Being A Connected View of the Principles of Evidence and the Method of Investigation, 2 Vol. London: Parker, 1843.
17. Laplace PS. Essai philosophique sur les probabilités, In : Oeuvres Complètes, vol VII. Paris: Gauthier-Villars, 1912: 25.
18. Howson C Urbach P. Scientific Reasoning: The Bayesian Approach, 2nd edition. Chicago and La Salle, Illinois: Open Court Publishing Company, 1993
19. Quine WO. Ontological Relativity and Other Essays. New York: Columbia Univ. Press, 1969: 23.