



❧ **THE FIRST PRINTED WORK IN PROBABILITY THEORY**



(HUYGENS) SCHOOTEN, Frans van. *Exercitationum mathematicarum libri quinque* [including: HUYGENS, Christian. *De ratiociniis in ludo aleae*]. Leiden, J. Elsevir, 1657.

Huygens treatise *De ratiociniis in aleae ludo* – the first printed work in probability theory – appears as an appendix to Schooten's works and, although brief, continued to form the best account of the subject until it was superceded by the more elaborate works of Bernoulli, Montmort, and De Moivre.

Here, Huygens solves the Problem of Points using the method of expectation. (Two players take turns flipping a p -biased coin and the first to accumulate N HEADS wins the ante. What would you pay to start with an advantage of K HEADS?)





BERNOULLI THEOREM –

❧ THE FUNDAMENTAL PRINCIPLE OF STATISTICS ❧

BERNOULLI, Jacob (1654-1705). *Ars Conjectandi, Opus Posthumum. Accedit Tractatus De Seriebus Infinitis, et Epistola Gallicè scripta de Ludo Pilae Reticularis* (Basel 1713).

The first systematic attempt to place the theory of probability on a firm basis and is still the foundation of much modern practice in all fields where probability is concerned – insurance, statistics and mathematical heredity tables.

Bernoulli proved that the fraction S/N of successes in repeated N trials approaches the probability of success, p , of each individual trial. Formally, for any ε and any δ , there is an N such that $P(|S/N - p| \leq \varepsilon) \geq 1 - \delta$, i.e., the sample mean is a consistent estimator of p .





❧ **THE ADVANTAGE OF TAKING THE MEAN** ❧

SIMPSON, Thomas (1710-1761). “On the advantage of taking the mean of a number of observations, in practical astronomy.” In *Philosophical Transactions* 49, part 1 (1755): 82-93. London: L. Davis & C. Reymers, 1756.

Simpson’s paper is a milestone in statistical inference. He was the first to attempt a mathematical proof of the law of large number i.e. that the mean result of several observations is nearer to the truth than any single observation. A key feature of the paper was that Simpson chose to focus “not on the observations themselves...but on the errors made in the observations, on the differences between the recorded observations and the actual position of the body being observed....[This] was the critical step that was to open the door to an applicable quantification of uncertainty” (Stigler 1986, 90-91). “Simpson was the first to characterize the errors in observations as independent events,” (Todhunter 1993) hence the first to use causal modeling.





❧ **BAYES THEOREM** ❧

BAYES, Thomas (1702-61). “An essay towards solving a problem in the doctrine of chances.” In *Philosophical Transactions* 53 London (1763): 370-418.

The first statement of Bayes’s Theorem for calculating “inverse probabilities,” which forms the basis for decision analysis, machine learning, and Bayesian networks. Bayesian or subjective decision theory is arguable the most comprehensive theory of decision-making; however, until the late 1980s, it had little application due to the stupefying complexity of the mathematics involved. The rapid advances in computation power and the development of key mathematical equations during the late 1980s and early 1990s made it possible to compute Bayesian networks with enough variables to be useful in practical applications (Christie 2004).





❧ **COMMON SENSE IN MATHEMATICAL LANGUAGE** ❧

LAPLACE, Pierre Simon (1749-1827). *Theorie Analytique des Probabilités*, Paris, 1814.

New to this 2nd edition is the introduction and a new concluding chapter 11. The introduction, published separately the same year under the title *Essai Philosophique sur les Probabilités*, expands upon the principles and applications set out in the original issue by detailing the practical applications of probability, theory in population studies, voting systems, the sex ratio, moral and political philosophy, games of chance, natural phenomena and miracles. The new chapter 11 deals with the probability of testimony. “The theory of probabilities, which Laplace described as *common sense expressed in mathematical language*, engaged his attention from its importance in physics and astronomy; and he applied his theory, not only to the ordinary problems of chances, but also to the inquiry into the causes of phenomena, vital statistics and future events” (Clerke).





❧ **CAUSALITY – AN EMPIRICAL CONCEPT** ❧

HUME, David (1711-1776). *A Treatise of Human Nature: Being an Attempt to Introduce the Experimental Method of Reasoning into Moral Subject* (London: printed for John Noon ... 1739).

The first attempt to apply Lock's empirical psychology to build a theory of knowledge, and from it to provide a critique of metaphysical ideas. Causation is a product of observations, and a learnable habit of the mind:

“Thus we remember to have seen that species of object we call FLAME, and to have felt that species of sensation we call HEAT. We likewise call to mind their constant conjunction in all past instances. Without any farther ceremony, we call the one CAUSE and the other EFFECT, and infer the existence of the one from that of the other” (page 156).





❧ **FALSIFICATION – THE KEY TO SCIENTIFIC INQUIRY** ❧

MILL, John Stuart (1806-1873). *A System of Logic, Ratiocinative and Inductive, Being a Connected View of the Principles of Evidence, and the Methods of Scientific Investigation ...* London, John W. Parker, 1843.

The *Logic* is `the first major installment of [Mill's] comprehensive restatement of an empiricist and utilitarian position ... [and] had a rapid and wide success. ... Mill held that the philosophy of experience was more likely than any other to encourage the development of society along liberal lines. He therefore held that it was a matter of considerable importance to show that empiricism was a viable alternative to the less progressive views – notably, Scottish common-sense philosophy and German idealism – which were then dominant. The *Logic* succeeded in doing this' (*The Encyclopedia of Philosophy*).





❧ **SYMBOLIC LOGIC** ❧

BOOLE, George (1815-64). *The Mathematical Analysis of Logic, Being an Essay towards a Calculus of Deductive Reasoning*. Cambridge: Macmillan..., 1847.

The first presentation of the Boolean algebra, the foundation of symbolic logic, and the basis for digital computing. This pamphlet precedes Boole's book-length presentation, *An Investigation of the Laws of Thought* (1853), and although brief, sets out his fundamental ideas: elective symbols and rules of operation on them the same as would hold in an algebra of the numbers 0 and 1.

“... it is a method resting upon the employment of Symbols, whose laws of combination are known and general, and whose results admit of a consistent interpretation ... It is upon the foundation of this general principle, that I propose to establish the Calculus of Logic, and that I claim for it a place among the acknowledged forms of mathematical analysis” (p. 4).





❧ **REGRESSION AND CORRELATION** ❧

GALTON, Francis (1822-1911). *Natural Inheritance*. London: Macmillan, 1889.

The first statistical study of biological variation and inheritance. Galton used his data as the basis for his “Law of filial regression,” by which he sought to demonstrate that long-established races habitually breed true. Galton concluded that evolution must take place by jumps and not minute steps... It was this kind of evidence that was to convince Bateson and others of the importance of mutations in forwarding the evolutionary process” (Forrest, *Galton*, p. 201).

Karl Pearson (1857-1936) confessed that it was this book which made him feel “like a buccaneer of `Drake’s days...’ and marginalize causation from statistics, in favor of regression and correlation” (Pearson 1934).





❧ **AUTOMATED REASONING** ❧

JEVONS, William Stanley (1835-1882). “On the mechanical performance of logical inference.” In *Philosophical Transactions* 160 London (1870): 497-518.

Jevon’s first published description of his logic machine, often called a “logical piano” because of its resemblance to that instrument. The “piano” was the first logic machine with enough power to solve complicated problems with superhuman speed. On the face of the piano are openings through which one can see letters representing the 16 possible combinations of four logical terms and their negatives. This action automatically eliminates from the face of the machine all combinations of terms which are inconsistent with the proposition just fed to the machine. After all premises have been fed to the device, its face is then examined to determine what conclusions can be drawn” (Lee, *Computer Pioneers*, 1995).





❧ **GOEDEL'S PROOF** ❧

GOEDEL, Kurt (1906-1978). "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I" (On Formally Undecidable Propositions of Principia Mathematica and related systems I). In *Monatshefte für Mathematik und Physik*, Vol 38 (Leipzig: 1931):173-98.

In this earthshaking paper, Goedel showed that it is impossible to establish the internal logical consistency of a very large class of deductive systems – elementary arithmetic, for example – unless one adopts principles of reasoning so complex that their internal consistency is as open to doubt as that of the systems themselves (Nagel and Newman 1960).





❧ **THE TURING TEST** ❧

TURING, Alan Mathison (1912-1954). “Computing machinery and intelligence.”
In *Mind* 59 (1950): 433-60.

FIRST EDITION. To answer the question of whether or not electronic computers think, Turing proposed a test: two parties, “X” and “Y”, engage in a conversation by teletype. Human X cannot know whether Y is a machine or a person. If X believes that Y is responding like a person after a specified period of time, and Y turns out to be a machine, then that machine may be defined as having the capacity to “think.”

“I believe that in about fifty years’ time it will be possible to programme computers with a storage capacity of about 10^9 , to make them play the imitation game so well that an average interrogator will not have more than 70 per cent chance of making the right identification after five minutes of questioning. The original question, ‘Can machines think?’ I believe to be too meaningless to deserve discussion” (p. 442).





❧ **COMPUTER CHESS** ❧

SHANNON, Claude E. (1916-2001). “A chess-playing machine.” In *Scientific American*, vol 182 No 2 (February 1950): 48-51.

In February 1950, a month before the appearance of his technical paper in the *Philosophical Magazine*, Shannon published this popular paper in the *Scientific American*, where he outlined the basic ideas behind computer chess in particular, and computer intelligence in general. Shannon does not embrace Turing’s test for thinking (published the same month), but reserves judgment: “Does a chess-playing machine of this type think? The answer depends entirely on how we define thinking”.





❧ **COMPUTER CHESS – MINIMAX SEARCH** ❧

SHANNON, Claude E. (1916-2001). “Programming a computer for playing chess.” In *Philosophical Magazine*, 7th series, 41, no. 314 (March 1950): 256-75.

The relevant history [of chess-playing programs] begins with a paper by Claude Shannon in 1949. He did not present a particular chess program, but discussed many of the basic problems involved. The framework he introduced has guided most of the subsequent analysis of the problem...” (Newell and Simon 1963, 42-44).

Here, Shannon introduces the ideas of minimax search and heuristic evaluation function. He was not aware though of alpha-beta pr.





❧ **MACHINE LEARNING** ❧

SAMUEL, Arthur Lee (1901-1990). “Some studies in machine learning using the game of checkers.” In *IBM Journal of Research and Development* 3, no. 3 (July 1959): 210-29 (plus folding table).

Samuel was the first to succeed in programming machines to “learn” from past errors, that is, to play a better game of checkers as time progresses. His checkers work also represents one of the earliest examples of non-numerical computation.





❧ **THE PERCEPTRON** ❧

ROSENBLATT, Frank (b. 1928) “The perceptron: A probabilistic model for information storage and organization in the brain.” In *Psychological Review* 65 (November 1958): 386-408.

The first precisely specified, computationally oriented neural network. The perceptron had the ability to recognize patterns and to associate new patterns with ones it had “seen” before on the basis of similarity; like the brain, it could function in the presence of “noise” and with damaged or missing connections. Later models were capable of trial-and-error learning and could be taught to emit ordered sequences of responses. Rosenblatt’s perceptron theory took a “connectionist” position with regard to the central questions of how information is stored or remembered, and how information contained in storage or memory influences recognition and behavior.

