






# The Theory That Would Not Die

*Reviewed by Andrew I. Dale*

the theory  that would  
 not die   
how bayes' rule cracked  
 the enigma code,  
hunted down russian  
submarines & emerged  
triumphant from two   
centuries of controversy  
sharon bertsch mcgrayne

---

**The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy**

Sharon Bertsch McGrayne

Yale University Press, April 2011

US\$27.50, 336 pages

ISBN-13: 978-03001-696-90

---

In the early 1730s Thomas Bayes (1701?–1761) was appointed minister at the Presbyterian Meeting House on Mount Zion, Tunbridge Wells, a town that had developed around the restorative chalybeate spring discovered there by Dudley, Lord North, in 1606. Apparently not one who was a particularly popular preacher, Bayes would be recalled today, if at all, merely as one of the minor clergy of eighteenth-century England, who also dabbled in mathematics. How is it, then, that Roger Farthing, author of an excellent history of Mount Zion, could describe Bayes as “to my mind, the greatest man to have lived in Tunbridge Wells” ([5, p. 167])? The answer is fairly simple: *Bayesian statistics*.

In 1763 Richard Price forwarded to the Royal Society of London an essay by Bayes [2] in which the following problem was addressed:

*Given* the number of times in which an unknown event has happened and failed: *Required* the chance that the probability of its happening in a single trial lies between any two degrees of probability that can be named.

---

Andrew I. Dale is emeritus professor of statistics at the University of KwaZulu-Natal in Durban, South Africa. His email address is da1e@ukzn.ac.za.

DOI: <http://dx.doi.org/10.1090/noti839>

The solution, given *more geometrico* as Proposition 10 in [2], can be written today in a somewhat compressed form as

posterior probability  $\propto$  likelihood  $\times$  prior probability.

Price himself added an appendix in which he used the proposition in a prospective sense to find the probability of the sun's rising tomorrow given that it has arisen daily a million times. Later use was made by Laplace, to whom one perhaps really owes modern Bayesian methods.

The degree to which one uses, or even supports, Bayes's Theorem (in some form or other) depends to a large extent on one's views on the nature of probability. Setting this point aside, one finds that the Theorem is generally used to update (to justify the updating of?) information in the light of new evidence as the latter is received, resulting in a strengthening of one's belief. Bayes's Theorem provides a codification of “learning from experience”, and *The Theory That Would Not Die* is concerned with the investigation and exposition of situations in which such learning has been both required and achieved.

A globally acceptable definition of *Bayesianism*, or *the Bayesian method*, seems hardly possible, for there are perhaps almost as many definitions as there are practitioners of the art. A useful and generally acceptable description is given as follows by Anthony O'Hagan: “The Bayesian method briefly comprises the following principal steps. *Likelihood ...Prior ...Posterior ...Inference*” ([8, p. 10]). It is also considered essential that the prior probability distributions be explicitly given (if these distributions are estimated from the data, the method, developed in the 1950s, is known as *Empirical Bayesianism*).

Laplace initially gave Bayes's Theorem in a simplified form as

$$\Pr[A_i|E] = \Pr[E|A_i] / \sum_1^n \Pr[E|A_j],$$

where, for instance,  $\Pr[E|A_i]$  is the probability of the event  $E$  given the cause  $A_i$ . This formula was developed in his later work to the continuous form one knows today, and he immediately applied it in his scientific investigations. A large study of births (more accurately baptisms) in France suggested to Laplace with strong probability that the chance of the birth of a boy exceeded that of a girl in Paris (and similarly in London).

Most statisticians these days are acquainted with the theoretical aspects of Bayesianism. Less well known, I suspect, are certain areas of practical application, and McGrayne's book is a "find" in this respect. So many, indeed, are the different applications considered that we cannot mention more than a few here.

Despite Laplace's work, his successors saw little virtue in Bayesian methods and almost succeeded in killing them off. One who kept the flickering flame alight was Joseph Louis François Bertrand with his work on projectiles for artillery field officers, and later, in the famous Dreyfus affair, the defence witness Henri Poincaré strongly supported the updating of the probability, in the light of later evidence, of the truth of a hypothesis as to whether a certain letter was a forgery.

Come the Great War, statisticians in the United States found themselves using Bayesian methods in connection with the making of decisions about injured workers and telephone communications. Yet these methods did not meet with general approval: "Early in the twentieth century" writes McGrayne, "theoreticians would change their attitudes toward Bayes' rule from tepid toleration to outright hostility" (p. 45)<sup>1</sup>.

Sporadic work continued after the war, and in 1925 Egon Sharpe Pearson published a long paper [9] in which he sought to put the theoretical rule to the test—"the most extensive exploration of Bayesian methods conducted between Laplace in the 1780s and the 1960s" (p. 49). Before the candle could be entirely guttered, the flame spluttered up in the work of Émile Borel, Frank Plump-ton Ramsey, and Bruno de Finetti and in the face of strong opposition from Ronald Aylmer Fisher and Jerzy Neyman. Only the sterling work of Harold Jeffreys—largely and undeservedly ignored by statisticians—at Cambridge kept the resuscitation on the go.

Pearson's idea ran as follows: let a sample of size  $n$  (say of taxicabs in London streets) be observed in which the number of taxis having registration

LX is  $p$  and that without is  $q$ . Then observe a further number  $m$  with the two numbers being  $r$  and  $s$ , respectively. Now repeat this experiment (with the same numbers  $n$  and  $m$ ) in different situations, and see whether "the distribution of observed values of  $r$  could be compared with the theoretical distributions which on Bayes' Theorem the knowledge of  $n$  and  $p$  should enable us to predict" ([9, p. 396]).

The damage done by German U-boats to Allied shipping during the Second World War made it imperative for the Allies to fathom the workings of the Enigma machines. These machines, used by the Germans to send encoded messages, became more and more sophisticated and complicated as the war progressed. Intensive research by a dedicated team under the leadership of Alan Turing at Bletchley Park in England eventually resulted in the cracking of the codes (the difficulty was considerably eased when a code book of encrypting tables was obtained from a sinking German submarine off Egypt). McGrayne writes "Turing was developing a homegrown Bayesian system. Finding the Enigma settings that had encoded a particular message was a classic problem in the inverse probability of causes" (p. 68).

During the war Turing paid a visit to the United States to discuss the work done at Bletchley Park with the U.S. Navy cryptographers. However "he tried in vain to explain the general principle that confirming inferences suggested by a hypothesis would make the hypothesis itself more probable" (pp. 75–6). Here Turing met Claude Shannon, who was working intensively on information theory. "Roughly speaking, if the posterior in a Bayesian equation is quite different from the prior, something has been learned; but when a posterior is basically the same as the prior guess, the information content is low" (p. 77). Turing's work was not altogether dismissed, however, for Eisenhower later said that "Bletchley Park's decoders had shortened the war in Europe by at least two years" (p. 81).

However, the success achieved by Bayesian methods during the war was not generally appreciated, it having been decided by those in authority that such results were to be kept classified. Bayes's Theorem was still seen as suspect: "'Bayes' still meant equal priors and did not yet mean making inferences, conclusions, or predictions based on updating observational data" (p. 87).

Part III of McGrayne's book records "the glorious revival". The first chapter here is devoted to the work of Arthur Bailey, who made extensive and important use of Bayes's Theorem in his work on credibility in casualty insurance as chief actuary at the New York State Insurance Department. His reason for so doing was made quite clear: "It

<sup>1</sup>References to page numbers without a citation are to the book being reviewed.

will be realized that all of the problems in which credibilities are used are problems in statistical estimation" ([1, p. 8]).

Chapter 8 details the entry of Bayesian methods into medical research, chiefly through the efforts, McGrayne claims, of Jerome Cornfield. The matter treated here is the connection of smoking with lung cancer and, later, heart disease. Small studies had been carried out before the Second World War, but the retrospective large study [3] published by Richard Doll and Austin Bradford Hill in 1950 gave firm (conclusive?) evidence of the connection between cigarette smoking and carcinoma of the lung. Doll and Hill, while identifying a seeming association, were careful to write that, "This is not necessarily to say that smoking causes carcinoma of the lung" ([3, p. 746]). In 1962, extending his important work, Cornfield studied the most critical risk factors for cardiovascular disease.

Bayesianism was later to be of further use in the medical context. For example, in the 1990s it was shown that mass screening of a large population for a rare disease (AIDS and HIV) would be unprofitable. As another example, while the test for the presence in men of high blood levels of prostate-specific antigen may be very efficient in identifying men who actually have prostate cancer, the disease is relatively so rare that in many, if not most, cases in which positive test results are returned the patient is found not to have cancer.

One of the most important situations in which Bayesianism scores heavily over frequentism is when there has been no past experience to draw on. (Here one might mention Good's paper [6] on the estimation of the population frequencies of species. Based on work done at Bletchley Park during the war, the paper was written in this vein to avoid Good's transgressing the Official Secrets Act). In Chapter 9 McGrayne looks at one important issue: the probability of a thermonuclear device exploding by mistake. While information was sparse, data were available on "the more dramatic incidents" (p. 122) involving nuclear weapons between 1950 and 1958. Results of such studies showed that there was certainly a positive probability of an accident, and some of the safety features suggested by the researchers Fred Charles Iklé and Albert Madansky were eventually implemented by the military.

While Bayesian and frequency methods sometimes result in the same conclusions, this is not always the case, and in [4] Edwards, Lindman, and Savage showed convincingly that in even moderate sample sizes Bayesian methods (with reasonable priors) and frequency  $p$ -values could lead to widely differing conclusions.

As the century progressed, Bayesian methods became more widely used. In the business field the prime developers were Howard Raiffa and Robert Osher Schlaifer. The emphasis was now switching from (merely) analyzing data to making decisions, and McGrayne notes that "some business Bayesians even dropped the prior odds called for by Bayes's rule" (p. 149). Whether such a move would be accepted as Bayesian might in fact be denied by many: Jeffreys, for instance, "was interested in making inferences from scientific evidence, not in using statistics to guide future action. To him, decision making ...was irrelevant" (p. 57).

Chapter 12 deals mainly with the investigation into the authorship of some papers published in *The Federalist* in 1787 and 1788. Twelve of these, published anonymously, were known to be by Alexander Hamilton or James Madison. McGrayne details the intensive undertaking by Mosteller and Wallace in the 1950s to attribute these papers to their correct authors, a task that took a decade's worth of work on word frequencies, the identification of "markers" (e.g., the presence of "while" or of "whilst") and Bayesian analysis. The prior odds were found, surprisingly, to be of little effect, and Mosteller and Wallace showed in [7] that, with "satisfyingly fat" odds, all twelve papers could be attributed to Madison.

John Tukey's use of Bayesian methods in psephology in the United States is dealt with in Chapter 13. While irritated by the "lack of methodology for quantifying Bayes' initial prior" (p. 168), Tukey (perhaps unstatedly) used a form of Empirical Bayesianism, with past election results being used to construct the prior. The analysis, carried out during several elections, presented forecasts of an election's final outcome as the results were coming in to NBC and turned out to be surprisingly accurate.

In the 1970s Norman Rasmussen undertook a study into the safety aspects of the nuclear power industry. A report issued in 1974 contained what was clearly a Bayesian investigation using probability distributions about equipment failure rates and human error. The report seemed to be of little effect: in 1979 the core of the Three Mile Island nuclear-generating Unit 2 was severely damaged. The Rasmussen report, McGrayne notes, now seemed prescient (p. 180).

Not all studies initially conceived as being ideally treated by Bayesian methods were in fact so eventually carried out. For example, on January 17, 1966, an American B-52 jet carrying four nonactivated bombs disintegrated in the air while refueling over the southeastern coast of Spain. While three of the bombs were found within twenty-four hours, the fourth proved more elusive.

John Craven, called in to handle the recovery operation, set up seven hypotheses (to which priors were assigned) as to what could have happened to the missing bomb—e.g., it had fallen free and would not be found in the debris of the plane; it had been carried far out to sea as a result of the deploying of both of the bomb's parachutes; a Spanish fisherman had seen the bomb fall into the sea.

Craven hired a team of mathematicians who soon discovered from their meetings with the military authorities that they had not been employed to use Bayes's Theorem and update their prior distribution. What was wanted was not an expectation that the bomb could be found: rather, "If the H-bomb was not found, the navy wanted to be able to prove statistically that it was not there" (p. 188). The prior probabilities of the hypotheses were never updated, despite data resulting from a search of the ocean floor from a ship. Fortunately optimal search methods eventually resulted in the desired finding of the bomb.

The final section of the book begins with a chapter in which a number of cases in which Bayesian methods have been used are fairly briefly discussed, cases in which the large amount of experimental data involved urged the development of computers and techniques (including efficient numerical integration procedures) that could handle such quantities of results. Long and difficult integrations, it was later realized, could be replaced by suitable sampling, an observation that led to the development of Markov chain Monte Carlo methods. Here a class of algorithms is taken for sampling from a probability distribution and based on the construction of a Markov chain having a desired equilibrium distribution. As a sample of the desired distribution, the state of the chain after a large number of steps is used.

In 1983 the U.S. Air Force sponsored a review of the estimates made by NASA of the probability of a shuttle failure. Using Bayesian methods, the researchers concluded that the odds on a failure were 1 in 35, NASA's estimate being 1 in 100,000. On January 28, 1986, the shuttle *Challenger* exploded on launch, a bitter vindication of Bayesianism.

The reader of *The Theory That Would Not Die* will realize that little is said here of prior distributions *per se*: most of the applications are concerned with the gathering of data and subsequent prognostication. Nevertheless, the effect of the prior generally diminishes with an increase in the amount of data.

McGrayne's giving the references in separate sections for each chapter is perhaps unfortunate. Sometimes things "fall through the cracks": for instance, Note 10 to Chapter 16 refers one to

"Smith (1984)", an item not listed in the bibliography to that chapter. Also, the system adopted makes it difficult to find a reference only vaguely remembered (was it in the list for Chapter N or Chapter M?).

For the student who is being exposed to Bayesian statistics for the first time, McGrayne's book provides a wealth of illustrations to whet his or her appetite for more. It will broaden and deepen the field of reference of the more experienced statistician, and the general reader will find an understandable, well-written, and fascinating account of a scientific field of great importance today.

## References

- [1] A. L. BAILEY, Credibility procedures: Laplace's generalization of Bayes' rule and the combination of collateral knowledge with observed data, *Proceedings of the Casualty Actuarial Society* 37 (1950), 7-23.
- [2] T. BAYES, An essay towards solving a problem in the doctrine of chances, *Philosophical Transactions of the Royal Society of London* 53 (published in 1764, in the volume for 1763), 370-418.
- [3] R. DOLL and A. BRADFORD HILL, Smoking and carcinoma of the lung, *British Medical Journal* 2, no. 4682, (1950), 739-748.
- [4] W. EDWARDS, H. LINDMAN and L. J. SAVAGE, Bayesian statistical inference for psychological research, *Psychological Review* 70, no. 3, (1963), 193-242.
- [5] R. FARTHING, *A History of Mount Sion, Tunbridge Wells, Phillimore, Chichester*, 2003.
- [6] I. J. GOOD, The population frequencies of species and the estimation of population parameters, *Biometrika* 40 (1953), 237-264.
- [7] F. MOSTELLER and D. L. WALLACE, *Inference and Disputed Authorship: The Federalist*, Addison-Wesley, Reading, 1964.
- [8] A. O'HAGAN, *Kendall's Advanced Theory of Statistics, Volume 2B, Bayesian Inference*, Edward Arnold, London, 1994.
- [9] E. S. PEARSON, Bayes' theorem, examined in the light of experimental sampling, *Biometrika* 17 (1925), 388-442.