REFERENCES

A Partially Annotated Bibliography

Abel, N. H. (1826), “Untersuchung der Functionen zweier unabhängig veränderlichen Grössen x und y, wie f(x, y), welche die Eigenschaft haben, dass f[z, f(x, y)] eine symmetrische Function von z, x und y ist.”, Jour. Reine u. angew. Math. (Crelle’s Jour.), 1, 11–15. First known instance of the associativity functional equation.

Abragam, A. (1961), Principles of Nuclear Magnetism, Oxford Science Publications, London. Apparently, he did not believe in phase coherences (as conveyed by off–diagonal elements of a density matrix) in quantum theory; so to predict the evolution of every new quantity, he was obliged to go into a new representation where that quantity is diagonal. But all this information was already present, independently of the representation, in a single density matrix with off–diagonal elements.


Akaike, Hirotugu (1980), “The Interpretation of Improper Prior Distributions as Limits of Data Dependent Proper Prior Distributions”, Jour. Roy. Stat. Soc. B42, pp. 46–52. A failed attempt to deal with the marginalization paradox, which never perceives that the paradox is just as present for proper priors as for improper ones. This has led others into much error and irrelevancy in dealing with time series. Discussed in Chapter 15.


Anscombe, F. J. (1963), “Sequential Medical Trials”, JASA 58, 365. Declares that the Sequential Analysis of Armitage (1960) is “a hoax”.


was written to refute a sentence in Jaynes (1965). We replied in “Comment on a review by P. W. Atkins”, Contemp. Phys. 28, pointing out some elementary facts of thermodynamics that had been well understood and correctly explained by Gibbs over 100 years before. In Jaynes (1992b) we go into much more detail on this.


Barnard, G. A. (1983), “Pivotal Inference and the Conditional View of Robustness (Why Have We for So Long Managed with Normality Assumptions?)”, in Box, Leonard & Wu (1983). Expresses somewhat the same surprise at the success of the normal distribution as Augustus de Morgan did 145 years earlier. We try to explain this in Chapter 7.


Barr, A. & Feigenbaum, E., editors (1981), The Handbook of Artificial Intelligence, 3 Vols., Wm. Kaufman, Inc., Los Altos CA. Contributions from over 100 authors. Volume 1 surveys search routines, one of the few aspects of AI that could be useful in scientific inference.


Bean, Wm. B. (1950), Aphorisms from the Bedside Teaching and Writings of Sir William Osler, (1849–1919). Henry Schumann, N. Y. Osler perceived the reasoning format of medical diagnosis in a form essentially identical with that given later by George Pólya. This was taken up by L. Lusted (1968) as the basis for Bayesian medical diagnosis computer programs.

Bell, E. T. (1937) Men and Mathematics, Dover Publications, Inc., N. Y. One needs to read this collection of biographical sketches because no substitute for it seems to exist; but let the reader be aware that Eric Temple Bell was also a well-known science fiction writer (under the pseudonym of John Taine) and this talent was not lost here. We can probably trust the accuracy of the names, dates, and documentable historical facts cited. But the interpretive statements tell us very little about the matter under discussion; they tell us a great deal about the fantasies and socio-political views of the writer, and the level of his comprehension of technical facts. For example (p. 167) he endorses, on the grounds of “social justice” the beheading of Lavoisier, the father of modern chemical nomenclature. He makes blatantly false accusations against Laplace, and equally falsely, portrays Boole as a saint who could do no wrong. Displays (p. 256) a ridiculous misconception of the nature of Einstein’s work, getting the sequence of facts backward. Tells us (p. 459) that Archimedes never cared for applications of mathematics!

Benford, F. (1938), "The law of anomalous numbers", Proc. Amer. Phil. Soc. 78, 551–572. Benford is probably the one referred to mysteriously by Warren Weaver (1963), p. 270. But, unknown to them, Simon Newcomb (1881) had noticed this phenomenon long before. See Raimi (1976) for many more details and references.


Bertrand, J. L. (1889), Calcul des probabilités, Gauthier–Villars, Paris. 2nd. edition, 1907, reprinted (1972) by Chelsea Publishing Co., New York. This work is usually cited only for the “Bertrand Paradox” which appears on pp. 4–5; but it is full of neat, concise mathematics as well as good conceptual insight, both of which are often superior to the presentations in recent works. He understood clearly how much our conclusions from given data must depend on prior information, an understanding that was lost in the later “orthodox” literature. We quote him on this at the end of Chapter 6. However, we disapprove of his criticism of the Herschel–Maxwell derivation of the Gaussian distribution that we give in Chap. 7; what he saw as a defect is what we consider its greatest merit, and a forerunner of Einstein’s reasoning. Well worth knowing and reviving today.


Bloomfield, P. (1976), *Fourier Analysis of Time Series: an Introduction*, J. Wiley & Sons, New York. Blackman–Tukey methods carried to absurd extremes (some 50 db below the noise level!!!) on alleged astronomical data on variable stars from Whittaker & Robinson (1924), which he fails to recognize as faked (he sees no difficulty in the implied claim that an unidentified observatory had clear skies on 600 successive midnights). As a result, the periodogram is giving zero information about variable stars; its top displays the two sine waves put into the simulated data, its bottom reveals only the spectrum of the digitizing errors. A potent demonstration of the folly of blind, unthinking application of a statistical procedure where it does not apply. A Bayesian would not be able to make such an error, because he would be obliged to think about his prior information concerning the phenomenon and the data taking procedure.


——— (1924), “Apropos of a Treatise on Probability”, review of Keynes (1921). Reprinted in Kyburg & Smokler (1981). Borel, like Bertrand (1889), understood very well how strongly probabilities must depend on our state of prior knowledge. It is a pity that neither undertook to demonstrate the important consequences of this in realistic applications; they might have averted fifty years of false teaching by others.


Boring, E. G. (1955), “The Present Status of Parapsychology”, Am. Scientist, 43, 108–116. Concludes that the curious phenomenon to be studied is the behavior of parapsychologists. Points out that, having observed any fact, attempts to prove that no natural explanation of it exists are logically impossible; one cannot prove a universal negative (quantum theorists who deny the existence of causal explanations please take note).


Bortkiewicz, L. V. (1898), *Das Gesetz der Kleinen Zahlen*, Teubner, Leipzig. Contains his famous fitting of the Poisson distribution to the number of German soldiers killed by the kick of a horse in successive years.

Boscovich, Roger J. (1770), *Voyage astronomique et geographique*, N. M. Tillard, Paris. Adjustment of data by the criterion that the sum of the corrections is zero, the sum of their magnitudes is made a minimum.

Box, G. E. P. & Tiao, G. C. (1973), *Bayesian Inference in Statistical Analysis*, Addison–Wesley, Reading MA. G. E. P. Box is, like L. J. Savage, a curious anomaly in this field; he was an assistant to R. A. Fisher and married his daughter, but became a Bayesian in issues of inference while remaining a Fisherian in matters of significance tests, which he held to be outside the ambit of Bayesian methods. In Jaynes (1985e) we argue that, on the contrary, any rational significance test requires the full Bayesian apparatus.


Bracewell, R. N. (1986), “Simulating the Sunspot Cycle,” *Nature*, 323, 516. Ronald Bracewell is perhaps the first author with the courage to present a definite prediction of future sunspot activity. We await the Sun’s verdict with interest.


Bross, I. D. J. (1963), “Linguistic Analysis of a Statistical Controversy”, Am. Stat. 17, 18. One of the most violent polemical denunciations of Bayesian methods in print – without the slightest attempt to examine the actual results they give! Should be read by all who want to understand why and by what means the progress of inference was held up for so long. Jaynes (1976) was written originally in 1963 as a reply to Bross, in circumstances explained in Jaynes (1983), p. 149.


proceeding to enough technical detail to be useful to practicing scientists. Be warned that what is called “Maximum Entropy” is in places distorted by ad hoc devices such as ‘windowing’ or ‘prefiltering’ the data—a practice that we condemn as destructive of some of the information in the data. Probability theory, correctly applied, is quite capable of extracting all the relevant information from the raw, unmutilated data and does best, with the least total computation, when it is allowed to do so freely.


Carnap, R. (1952), The Continuum of Inductive Methods, University of Chicago Press.

Cheeseman, P. (1988), “An Inquiry into Computer Understanding”, Comput. Intell. 4, 58–66. See also the following 76 pages of discussion. This attempt to explain Bayesian principles to the Artificial Intelligence community ran into incredible opposition, from discussants who had no comprehension of what he was doing. The situation is described in Jaynes (1990b).


When first issued, this work was described as “the only textbook on statistics that is not twenty years behind the times”. It is now more than thirty years behind the times, because they could not accept the notion of a probability that is not a frequency, and so did not appreciate the fact that a straight Bayesian approach leads to all the same results with an order of magnitude less formal machinery. Still, it is an interesting and entertaining exposition of Wald’s original ideas, far easier to read than Wald (1950).


Cobb, L. & Watson, B. (1980), “Statistical Catastrophe Theory: An Overview”, Mathematical Modelling, 1, pp. 311–317. We have no quarrel with this work, but wish to add two historical footnotes. (1) Their “stochastic differential equation” is what physicists have called a “Fokker–Planck equation” since about 1917. However, we are used to having our statistical work attributed to Kolmogorov by mathematicians. (2) Stability considerations of multiple–valued “folded” functions of the kind associated today with the name of René Thom are equivalent to convexity properties of a single–valued entropy function, and these were given by J. Willard Gibbs in 1873.


Cook, A. (1994), The Observational Foundations of Physics, Cambridge University Press, U. K. Notes that physical quantities are defined in terms of the experimental arrangement used to measure them. Of course, this is just what Niels Bohr emphasized in 1927.


Cournot, A. A. (1843), Exposition de la theorie des chances et des probabilities; L. Hachette, Paris. Reprinted (1984) in Oeuvres complètes, J. Vrin, Paris. One of the first of the attacks against Laplace, which were carried on by Ellis, Boole, Venn, E. T. Bell, and others to this day.

Cox, D. R. & Hinkley, D. V. (1974), Theoretical Statistics, Chapman & Hall, London. Reprints 1979, 1982. Mostly a repetition of old sampling theory methods, in a bizarre notation that can make the simplest equation unreadable. However, it has many useful historical summaries and side remarks noting limitations or extensions of the theory, that cannot be found elsewhere. Bayesian methods are introduced only in the penultimate Chapter 10; and then the authors proceed to repeat all the old, erroneous objections to them, showing no comprehension that these were ancient misunderstandings long since corrected by Jeffreys (1939), Savage (1954), and Lindley (1965). One prominent statistician, noting this, opined that Cox & Hinkley had “set statistics back 25 years.”


Cox, R. T. (1946), “Probability, Frequency, and Reasonable Expectation”, Am. Jour. Phys. 14, 1–13. In our view, this article was the most important advance in the conceptual (as opposed to the purely mathematical) formulation of probability theory since Laplace.


Cramér, H. (1946), Mathematical Methods of Statistics, Princeton University Press. This marks the heyday of supreme confidence in confidence intervals over Bayesian methods, asserted as usual on purely ideological grounds, taking no note of the actual results that the two methods yield. Comments on it are in Jaynes (1976, 1986a).


Crick, Francis (1988), What Mad Pursuit, Basic Books, Inc., New York. A reminiscence of his life and work, full of important observations and advice about the conduct of science in general, and fascinating technical details about his decisively important work in biology – most of which occurred several years after the famous Crick – Watson discovery of the DNA structure. Almost
equally important, this is an antidote to Watson (1968); we have here the other side of the DNA Double Helix story as Crick recorded it in 1974, with a different recollection of events. From our viewpoint, this work is valuable as a case history of important scientific discoveries made without help of probabilistic inference in our mathematical form, but – at least in Crick’s mind – obeying its principles strictly, in the qualitative form given by Pólya. We wish that theoretical physicists reasoned as well.


Czuber, E. (1908), *Wahrscheinlichkeitsrechnung und Ihre Anwendung auf Fehlerausgleichung*, Teubner, Berlin; 2 Vols. Some of Wolf’s famous dice data may be found here.


Daniell, G. J. & Potton, J. A. (1989), “Liquid Structure Factor Determination by Neutron Scattering – Some Dangers of Maximum Entropy”, in Skilling (1989), pp. 151 – 162. The “danger” here is that a beginner’s first attempt to use maximum entropy on a complex problem may be unsatisfactory because it is answering a different question than what the user had in mind. So the first effort is really a “training exercise” which makes one aware of how to formulate the problem properly.


David, F. N. (1962), *Games, Gods and Gambling*, Griffin, London. A history of the earliest beginnings of probability theory. Notes that in Archaeology, “the farther back one goes, the more fragmentary is the evidence.” Just the kind of deep insight that we could find nowhere else.


Dawkins, R. (1987), *The Blind Watchmaker*, W. W. Norton & Co., New York. An answer to the unceasing attacks on Darwin’s theory, by religious fundamentalists who do not understand what Darwin’s theory is. Richard Dawkins, Professor of Zoology at Oxford University, goes patiently into much detail to explain, as did Charles Darwin 120 years earlier, why the facts of Nature can be accounted for as the operation of Natural Law, with no need to invoke teleological purpose; and we agree entirely. Unfortunately, Dawkins’ enthusiasm seem to outrun his logic; on the cover he claims that it also explains a very different thing: “Why the evidence of evolution reveals a universe without design”. We do not see how any evidence could possibly do this; elementary logic warns us of the difficulty of proving a negative.

Dawkins’ struggle against fundamentalist religion has continued; in 1993 the Starbridge Lectureship of Theology and Natural Science was established in the Faculty of Divinity of Cambridge University. Dawkins wrote in the national press to deplore this and stress the vacuity of theology
contrasted with the value of science. This prompted the Cambridge Nobel Laureate chemist Max Perutz to issue an unperceptive rejoinder, saying: “Science teaches us the laws of nature, but religion commands us how we should live. ... Dr. Dawkins does a disservice to the public perception of scientists by picturing them as the demolition squad of religious beliefs.” It appears to us that Dawkins was deploring arbitrary systems of theology, rather than ethical teachings; again, these are very different things.


_______ (1974b), Theory of Probability, 2 Vols. J. Wiley & Sons, Inc., New York. Adrian Smith’s English translation could not hide the wit and humor of this work. Bruno de Finetti was having great fun writing it; but he could scarcely write two sentences without injecting some parenthetic remark about a different topic, that suddenly popped into his mind, and this is followed faithfully in the translation. Full of interesting information that all serious students of the field ought to know; but impossible to summarize, because of its chaotic disorganization. Discussion of any one topic may be scattered over a half-dozen different Chapters without cross-references, so one may as well read the pages at random.


de Groot, M. H. (1975), Probability and Statistics, Addison-Wesley Publishing Co., Reading MA; 2nd edition (1986). This textbook is full of useful results, but represents an intermediate transitional phase between orthodox statistics and modern Bayesian inference. Morrie de Groot (1931-1989), a Ph.D. student of the transitional Bayesian L. J. Savage, saw clearly the technical superiority of Bayesian methods and was a regular attendant and speaker at our twice-yearly NSF-NBER Bayesian Seminars; but he still retained the terminology, notation, and general absolutist mindset of orthodoxy. Thus he still speaks of ‘true probabilities’ and ‘estimated probabilities’ as if the former had a real existence, and distinguishes sharply between ‘probability theory’ and ‘statistical inference’ as if they were different topics. This does not prevent him from obtaining the standard useful results, often by continuing the orthodox habit of inventing ad hoc devices instead of application of the rules of probability theory. [Our present relativist theory recognizes that there is no such thing as an ‘absolute’ probability, because all probabilities express, and are necessarily conditional on, the user’s state of information. This makes the general principles applicable uniformly to all problems of inference, with no need for ad hoceries.] A biography and bibliography of Morris de Groot may be found in Statistical Science, vol 6, pp. 4–14 (1991).


_______ (1872), *A Budget of Paradoxes*, 2 Vols. Sophia de Morgan, editor, London. 2nd edition, D. E. Smith, editor (1915); reprinted as one Volume by Dover Publications, Inc. (1954). Augustus de Morgan (1806–1871) was a mathematician and logician, at University College, London from 1828–1866. He collected notes concerning not only logic, but anomalies of logic; the latter are preserved in this delightful account of the activities of circle-squarers, anti-Copernicans, anti-Newtonians, religious fanatics, numerologists, and other demented souls that abounded in 19th Century England. It gives a vivid picture of the difficulties that serious scholars had to overcome in order to make any forward progress in science. An inexhaustible supply of amusing anecdotes.


Denbigh, K. G. & Denbigh, J. S. (1985), *Entropy in Relation to Complete Knowledge*, Cambridge University Press. This is perhaps the first example of an entire book written for the purpose of attacking a single sentence in a tutorial paper. In Jaynes (1965) we noted (as had L. Boltzmann, G. N. Lewis, Arthur Eddington, J. von Neumann, and Eugene Wigner before us) that “entropy is an anthropomorphic concept”. The meaning was that it is not only a measure of phase volume compatible with a macrostate; it is also a measure of human ignorance as to the microstate when we know only the macrostate. That is, it indicates the number of bits of additional information we would need in order to locate a macroscopic thermodynamic system in a definite microstate; this is not an opinion, but a theorem. Although the above authors had been expounding this viewpoint for decades without incurring any criticism for it, as soon as we said the same thing, for reasons we cannot understand, this caused an explosion in the mind of Kenneth Denbigh, who proceeded to issue violent denunciations of our view. It seems to us that his arguments are self-refuting and do not call for any reply. But the issue was taken up in turn by Atkins (1986), to which we were finally moved to reply.


Dyson, F. J. (1979), *Disturbing the Universe*, Harper & Row, Publishers, New York. A collection of personal reminiscences and speculations extending over some fifty years. Ninety percent of it is irrelevant to our present purpose; but one must persist here, because Freeman Dyson played a very important part in the development of theoretical physics in the mid Twentieth Century. His reminiscences about this are uniquely valuable, but unfortunately scattered in small pieces over several Chapters. Unlike some of his less thoughtful colleagues, Dyson saw correctly many fundamental things about probability theory and quantum theory (but in our view missed some others equally fundamental). Reading this work is rather like reading Kepler, and trying to extract the tiny nuggets of important truth.

Edwards, A. W. F. (1972), *Likelihood*, Cambridge University Press. Anthony Edwards was the last student of R. A. Fisher; and although he understands all the technical facts pertaining to Bayesian methods as well as anybody, some mental block prevents him, as it did Fisher, from accepting their obvious consequences. So we must, sadly, part company and proceed with the constructive development of inference without him.


Efron, B. & G. Gong (1983), “A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation”, Am. Stat. 37, pp. 36–48. Orthodox statisticians have continued trying to deal with problems of inference by inventing arbitrary *ad hoc* procedures instead of applying probability theory. Three recent examples are explained and advocated here. Of course, they all violate our desiderata of rationality and consistency; the reader will find it interesting and instructive to demonstrate this and compare their results with those of the Bayesian alternatives.


Feinberg, S. E. & Hinkley, D. V. (1990), *R. A. Fisher: An Appreciation*, Lecture Notes in Statistics #1, Springer–Verlag, Berlin. This is the second printing of the work, which appeared originally in 1979. A valuable source, if it is regarded as an historical document rather than an account
of present statistical principles. Rich in technical details of his most important derivations and gives a large bibliography of his works, including four books and 294 published articles. But in its adulation of Fisher it fails repeatedly to note something that was already well established in 1979: the simpler and unified methods of Jeffreys, which Fisher rejected vehemently, actually accomplished everything that Fisher’s methods did, with the same or better results and almost always more easily. In addition, they deal easily with technical difficulties (such as nuisance parameters or lack of sufficient statistics) which Fisher was never able to overcome. Thus this work tends also to perpetuate harmful myths.


Fieller, E. C. (1954), “Some problems in interval estimation”, J. Roy Stat. Soc. B 16, 175–185. This and the contiguous paper by Creasy (1954) became famous as ‘The Fieller–Creasy Problem’ of estimating the ratio $\mu_1/\mu_2$ of means of two normal sampling distributions. It generated a vast amount of discussion and controversy because orthodox methods had no principles for dealing with it – and for decades nobody would deign to examine the Bayesian solution. It is a prime example of an estimation problem, easily stated, for which only Bayesian methods provide the technical apparatus required to solve it. It is finally considered from a Bayesian standpoint by José Bernardo (1977). For us, it is a straightforward exercise for the reader in our Chapter on Estimation with a Gaussian distribution.


Fischer, E. P. & Lipson, C. (1988), *Thinking About Science: Max Delbrück and the Origins of Molecular Biology*, Norton, New York. For some time we have seen Max Delbrück referred to as “one of Niels Bohr’s greatest and most successful students”. It is true that he has played a very important role in the modern development of biology as the leader of the “phage school”; yet what has emerged was nearly the opposite of his intentions. As he himself has noted, his original goal – to inculcate the ideas of the Copenhagen interpretation of quantum theory into biology and to learn new principles of physics from biology – has not been realized, all the new developments involving definite, reliable mechanisms that would be understood at once in a machine shop. The role of “quantum effects” in biology seems limited to their role in all of chemistry: to account for the binding energies – hence the stability – of molecules. The uncertainty principle has as yet found no functional role at all in biology, nor have any new physical principles emerged; and we predict with confidence that this will continue to be true.


——— (1915), “Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population”, Biometrika, 10, pp. 507–521.


(1930b), *The Genetical Theory of Natural Selection*, Oxford University Press. Second revised Edition by Dover Publications, Inc., New York (1958). Here Fisher shows that Mendelian genetics is not in conflict with Darwinian evolution theory, as Mendelians supposed in the early 20th Century; on the contrary, the ‘particulate’ or ‘discrete’ nature of Mendelian inheritance clears up some outstanding difficulties with Darwin’s theory, resulting from the assumption of blending inheritance which most biologists – including Darwin himself – took for granted in the 1860’s. Recall that Mendel’s work, with its lore of dominant and recessive genes, etc., was later than Darwin’s; but Darwin (1809 – 1882) never knew of it and it was not generally known until after 1900. The reinterpretation of Darwin’s theory in these terms, by Fisher and others, is now known as Neo-Darwinism. By the time of Fisher’s second (1958) edition the existence of mutations caused by radioactivity was well established, those caused by failures of DNA replication had become highly plausible, and genetic recombination (which had been suggested by August Weismann as early as 1886) was recognized as still another mechanism to provide the individual variations on which Natural Selection feeds, but whose origin was puzzling to Darwin. So Fisher added many new paragraphs, in smaller type, pointing out this newer understanding and its implications; how Darwin would have enjoyed seeing these beautiful solutions to his problems! Fisher’s real, permanent contributions to science are in works like this, not in his statistical teachings, which were an advance in the 1920’s, but have been a retarding force since the 1939 work of Jeffreys.


(1938), *Statistical Tables for Biological, Agricultural and Medical Research* (with F. Yates), Oliver & Boyd, Edinburgh; five later editions to 1963.


(1956), *Statistical Methods and Scientific Inference*, Oliver & Boyd, London. Second Revised Edition, Hafer Publishing Co., New York, 1959. Fisher’s final book on statistics, in which he tries to sum up his views of the logical nature of uncertain inference. One discerns a considerable shift of position from his earlier works – even admitting, occasionally, that he had been wrong before. He is now more sympathetic toward the role of prior information, saying that recognizable subsets should be taken into account and that prior ignorance is essential for the validity of fiducial estimation. He shows his old power of intuitive insight in his neat explanation of Gödel’s theorem, but also some apparent lapses of memory and numerical errors. Every serious student of the subject should read this work slowly and carefully at least twice, because the depth of thinking is so great that his meaning will not be grasped fully on a single reading. Also, Fisher goes into several specialized topics that we do not discuss in the present work.


Fraser, D. A. S. (1980), Comments on a paper by B. Hill, in *Bayesian Statistics*, J. M. Bernardo *et al.*, editors, University Press, Valencia, Spain, pp. 56–58. Claims to have a counter-example to the likelihood principle. But it is the same as the tetrahedron problem discussed in Chapter 15 above; the correct solution to that problem was not known in 1980.


Galton, F. (1863), *Meteorographica*, London; MacMillan. Here this remarkable man invents weather maps and from studying them discovers the “anticyclone” circulation patterns in the northern hemisphere.


Gardner, M. (1957), *Fads and Fallacies in the Name of Science*, Dover Publications, Inc. A kind of 20th Century sequel to de Morgan (1872), with attention directed more to fakers in science than to their colleagues in mathematics. Here we meet both the sincere but tragically misguided souls, and the deliberate frauds out to make a dishonest dollar from the gullible.

Gardner, M. (1981), *Science – Good, Bad, and Bogus*, Paperbound edition (1989), Prometheus Books, Buffalo N. Y. A sequel to the previous work, with a sobering message that everyone ought to note. Particular details on several recent trends; the Creationist who utilizes TV to carry attacks on Darwin’s theory to millions, while grossly misrepresenting what Darwin’s theory is; the ESP advocate who invades scientific meetings to try to invoke Quantum Theory in his support, although he has no comprehension of what Quantum Theory is; the Gee Whiz publicist who turns every tiny advance in knowledge (artificial intelligence, chaos, catastrophe theory, fractals) into a revolutionary crusader cult; the professional Disaster Monger who seeks personal publicity through inventing ever more ridiculous dangers out of every activity of Man; and most frightening of all, the eagerness with which the news media give instant support and free publicity to all this. Today, our airwaves are saturated with bogus science and medieval superstitions belittling and misrepresenting real, responsible science. In the Introduction, Gardner documents the indignant refusal of network executives to correct this, on grounds of its profitability. Then at what point does persistent, deliberate abuse of freedom of speech for profit become a clear and present danger to society? See also Rothman (1989); Huber (1992).


Gauss, K. F. (1823), *Theoria combinationis observationum erroribus minimis obnoxiae; also Supplementum*, 1826; Dieterich, Göttingen.


Gell–Mann, M. (1992) “Nature Conformable to Herself”, Bulletin of the Santa Fe Institute, 7, pp. 7–10. Some comments on the relation between mathematics and physics; this Nobel Laureate theoretical physicist is, like us, happy that the ‘plague of Bourbakism’ is finally disappearing, raising the hope that mathematics and theoretical physics may become once more mutually helpful partners instead of adversaries.


Gnedenko, B. V. & Kolmogorov, A. N. (1954), Limit Distributions for Sums of Independent Random Variables, Addison-Wesley, Cambridge MA. On p. 1 we find the curious statement: “In fact, all epistemologic value of the theory of probability is based on this: that large-scale random phenomena in their collective action create strict, non-random regularity.” This was thought by some to serve a political purpose in the old USSR; in any event, the most valuable applications of probability theory today are concerned with incomplete information and have nothing to do with those so-called ‘random phenomena’ which are still undefined in theory and unidentified in Nature.


Goldman, S. (1953), Information Theory, Prentice-Hall, Inc., New York. We would like to put in a friendly plug for this work, even though it has a weird reputation in the field. The author, in recounting the work of Norbert Wiener and Claude Shannon, explains it for the benefit of beginners much more clearly than Wiener did, and somewhat more clearly than Shannon. Its weirdness is the result of two unfortunate accidents: (1) a misspelled word in the title of Chapter 1 escaped both the author and the publisher, providing material for dozens of cruel jokes circulating in the 1950’s; (2) on p. 295 there is a photograph of Gibbs, with the caption: “J. Willard Gibbs (1839–1903), whose ergodic hypothesis is the forerunner of fundamental ideas in information theory.” Since Gibbs never mentioned ergodicity, this is a source of more jokes. However, the author is guilty only of trusting the veracity of Wiener (1948).

whose importance is out of all proportion to its small size. Still required reading for every student of Scientific Inference; and can be read in one evening.

——— (1965), *The Estimation of Probabilities*, Research Monographs #30, MIT Press, Cambridge, MA. Jack Good persisted in believing in the existence of ‘physical probabilities’ that have some kind of reality independently of human information; hence the (to us) incongruous title.


——— (1983), *Good Thinking*, University of Minnesota Press. Reprints of 23 articles, scattered over many topics and many years, plus a long bibliography of other works. There are about 2000 short articles like these by Good, found throughout the statistical and philosophical literature starting in 1940. Workers in the field generally granted that every idea in modern statistics can be found expressed by him in one or more of these articles; but their sheer number made it impossible to find or cite them, and most are only one or two pages long, dashed off in an hour and never developed further. So for many years, whatever one did in Bayesian statistics, one just conceded priority to Jack Good by default, without attempting the literature search for the relevant article, which would have required days. Finally, this book provided a bibliography of most of the first 1517 of these articles (presumably in the order of their writing, which is not the order of publication) with a long index, so it is now possible to give proper acknowledgment of his works up to 1983. Be sure to read Chapter 15, where he points out specific, quantitative errors in Karl Popper’s work and demonstrates that Bayesian methods, which Popper rejects, actually correct those errors.

Gould, Stephen Jay (1989), *Wonderful Life: The Burgess Shale and the Nature of History*, W. W. Norton & Co., New York. A tiny region in the Canadian Rockies had exactly the right geological history so that soft-bodied animals were preserved almost perfectly. As a result we now know that the variety of life existing in early Cambrian time was vastly greater than had been supposed; this has profound implications for our view of evolution. Gould seems fanatical in his insistence that ‘evolution’ is not synonymous with ‘progress’. Of course, anyone familiar with the principles of physics and chemistry will agree at once that a process that proceeded in one direction can also proceed in the opposite one. Nevertheless, it seems to us that at least 99% of observed evolutionary change *has in fact* been in the direction of progress (more competent, adaptable creatures). We also think that Darwinian theory, properly stated in terms conforming to present basic knowledge and present Bayesian principles of reasoning, predicts just this.


attempted to make a survey of Ireland many years before Halley, but did not reason carefully enough to produce a meaningful result. Greenwood ends in utter confusion over whether Petty is or is not the real author of Graunt’s book, apparently unaware that Petty’s connection is that he edited the fifth (posthumous) edition of Graunt’s work; and it was Petty’s edition that Halley referred to and saw how to correct. All this had been explained long before, with amusing sarcasm, by Augustus de Morgan (1872, I, 113–115).


Haldane, J. B. S. (1957), “Karl Pearson, 1857–1957”, *Biometrika* 44, 303–313. Haldane’s writings, whatever the ostensible topic, often turned into political indoctrination for socialism. In this case it made some sense, since Karl Pearson was himself a political radical. Haldane suggests that he may have changed the spelling of his name from ‘Carl’ to ‘Karl’ in honor of Karl Marx, and from this Centenary oration we learn that V. I. Lenin quoted approvingly from Karl Pearson. Haldane was Professor of Genetics at University College, London in the 1930’s, but he resigned and moved to India as a protest at the failure of the authorities to provide the financial support he felt his Department needed. It is easy to imagine that this was precisely what those authorities, exasperated at his preoccupation with left-wing politics instead of genetics, hoped to bring about. An interesting coincidence is that Haldane’s sister, Naomi Haldane Mitchison, married a Labour MP and carried on the left-wing cause. James D. Watson was a guest at her home at Christmas 1951, about a year before discovering the DNA helix structure. He was so charmed by the experience that his 1968 book, “The Double Helix”, is inscribed: “For Naomi Mitchison.”


Howson, C. & Urbach, P. (1989), *Scientific Reasoning: The Bayesian Approach*, Open Court Publishing Co., La Salle, Illinois. A curiously outdated work, which might have served a useful purpose 60 years earlier. Mostly a rehash of all the false starts of philosophers in the past, while offering no new insight into them and ignoring the modern developments by scientists, engineers, and economists which have made them obsolete. What little positive Bayesian material there is, represents a level of understanding that Harold Jeffreys had surpassed 50 years earlier, minus the mathematics needed to apply it. They persist in the pre-Jeffreys notation which fails to indicate the prior information in a probability symbol, take no note of nuisance parameters, and solve no problems.


Huber, P. (1992), *Galileo’s Revenge: Junk Science in the Courtroom*, Basic Books, Inc., N. Y. Documents the devastating effects now being produced by charlatans and crackpots posing as scientists. They are paid to give ‘expert’ testimony that claims all sorts of weird causal relations that do not exist, in support of lawsuits that waste billions of dollars for consumers and businesses. The phenomena of pro-causal and anti-causal bias are discussed in Chapters 5, 16,
17. At present we seem to have no effective way to counteract this; as noted by Gardner (1981), the News Media will always raise a great wind of publicity, giving support and encouragement to the charlatans while denying responsible scientists a hearing to present the real facts. It appears that the issue of what is and what is not valid scientific inference must soon move out of Academia and become a matter of legislation – a prospect even more frightening than the present abuses.


Jansson, P. A., editor (1984), *Deconvolution, with Applications in Spectroscopy*, Academic Press, Inc., Orlando FL. Articles by nine authors, summarizing the State of the Art (mostly linear processing) as it existed just before the introduction of Bayesian and Maximum Entropy methods.


——— (1963d), Review of *Noise and Fluctuations*, by D. K. C. MacDonald, Am. Jour. Phys 31, 946. Cited in Jaynes (1976) in response to a charge by Oscar Kempthorne that physicists have paid little attention to noise; notes that there is no area of physics in which the phenomenon of noise does not present itself. As a result, physicists were actively studying noise and knew the proper way to deal with it, long before there was any such thing as a statistician.


——— (1985b), “Entropy and Search Theory”, in Smith & Grandy (1985), pp. 443–454. Shows that the failure of previous efforts to find a connection between information theory and search theory were due to use of the wrong entropy expression. In fact, there is a very simple and general connection, as soon as we define entropy on the deepest hypothesis space.


——— (1986a), “Bayesian Methods: General Background”, in Justice (1986). A general, non–technical introductory tutorial for beginners, intended to explain the terminology and viewpoint, and warn of common pitfalls of misunderstanding and communication difficulties.


——— (1986c), “Some Applications and Extensions of the de Finetti Representation Theorem”, in Goel & Zellner, (1986); pp. 31–42. The theorem, commonly held to apply only to infinite
exchangeable sequences, remains valid for finite ones if one drops the non-negativity condition on the generating function. This makes it applicable to a much wider class of problems.


_________ (1992b), “The Gibbs Paradox”, in Proceedings of the 11’th Annual MAXENT Workshop, Maximum Entropy and Bayesian Methods, Seattle, 1991, C. R. Smith & G. Ericksen, editors, Kluwer Academic Publishers, Holland; pp. 1–21. There is no paradox; Gibbs explained it all in his early work on Heterogeneous Equilibrium, but this was missed by later readers who examined only his Statistical Mechanics. The range of valid applications of classical thermodynamics is far greater for one who understands this. Note a misprint; the text equation preceding Eq. (14) should be: \( f(1) = \log(\frac{Ck^2}{2}) \).

261–275. A response to the contributors to this Festschrift volume marking the writer’s 70’th birthday, with 22 articles by my former students and colleagues.


Jeffreys, Wm. H. (1990), “Bayesian Analysis of Random Event Generator Data”, Jour. Scientific Exploration, 4, pp. 153-169. Shows that orthodox significance tests can grossly overestimate the significance of ESP data; Bayesian tests yield defensible conclusions because they do not depend on the intentions of the investigator.

Jeffrey, R. C. (1983), The Logic of Decision, 2nd edition, Univ. of Chicago Press. Attempts to modify Bayes’ theorem in an ad hoc way; as discussed in Chapter 5, this necessarily violates one of our desiderata.


Jeffreys, Lady Bertha Swirles (1992) “Harold Jeffreys from 1891 to 1940”, Notes Rec. R. Soc. Lond. 46, 301–308. A short, and puzzlingly incomplete, account of the early life of Sir Harold Jeffreys, with a photograph of him in his 30’s. Detailed account of his interest in botany and early honors (he entered St. John’s College, Cambridge as an undergraduate, in 1910; and that same year received the Adams memorial prize for an essay on ‘Precession and Nutation’). But, astonishingly, there is no mention at all of his work in probability theory! In the period 1919–1939 this resulted in many published articles and two books (Jeffreys, 1931, 1939) of very great importance to scientists today. It is, furthermore, of fundamental importance and will remain so long after all his other work recedes into history. Bertha Swirles Jeffreys was also a physicist, who studied with Max Born in Göttingen in the late 1920’s and later became Mistress of Girton College, Cambridge.


Johnson, L. T. (1996), “The Real Jesus” A contribution to the current renewed controversies over the ‘historical Jesus’ that have surfaced periodically since the time of Laplace. The author is very conscientious in separating what is based on historical reality and what is based on faith; nevertheless, he persists in clinging to his faith, accepting miracles literally while ignoring the evidence of science. In this respect the work is a backward step from Conybeare (1958), violating the principles of consistent reasoning, which demand that all the relevant evidence be taken into account.


Kendall, M. G. & Moran, P. A. P. (1963), Geometrical Probability, Griffin, London. Much useful mathematical material, all of which is readily adapted to Bayesian pursuits.


Kendall, M. G. & Stuart, A. (1961), The Advanced Theory of Statistics: Volume 2, Inference and Relationship, Hafner Publishing Co., New York. This represents the beginning of the end for the confidence interval; while they continued to endorse it on grounds of “objectivity”, they noted so many resulting absurdities that readers of this work were afraid to use confidence intervals thereafter. In Jaynes (1976) we explained the source of the difficulty and showed that these absurd results are corrected automatically by use of Bayesian methods.


Khinchin, A. I. (1957), *Mathematical Foundations of Information Theory*, Dover Publications, Inc. Attempts of a mathematician to ‘rigorize’ Shannon’s work. But we do not think it was in need of this. In any event, when one tries to work directly on infinite sets from the beginning, the resulting theorems just do not refer to anything in the real world. Khinchin was probably careful enough to avoid actual error and thus produced theorems valid in his imaginary world; but we note in Chapter 15 some of the horrors that have been produced by others who tried to do mathematics this way.


Kurtz, P. (1985), *A Skeptic’s Handbook of Parapsychology*, Prometheus Books, Buffalo N. Y. Several Chapters have relevant material; see particularly Chapter 11 by Betty Markwick.


——— (1783), *Histoire d’l’Académie*, pp. 423–467. An early exposition of the properties of the “Gaussian” distribution. Suggests that it is so important that it should be tabulated.


Lewis, G. N. (1930) “The Symmetry of Time in Physics”, *Science*, **71**, 569. An early recognition of the connection between entropy and information, showing an understanding far superior to what many others were publishing 50 years later.
Lighthill, M. J. (1957), *Introduction to Fourier Analysis and Generalised Functions*, Cambridge Univ. Press. Required reading for all who have been taught to mistrust delta-functions. See the review by Freeman Dyson (1958). Lighthill and Dyson were classmates in G. H. Hardy’s famous course in ‘Pure Mathematics’ at Cambridge University, at a time when Fourier analysis was mostly preoccupied with convergence theory, as in Titchmarsh (1937). Now with a redefinition of the term ‘function’ as explained in our Appendix B, all that becomes nearly irrelevant. Dyson states that Lighthill ‘lays Hardy’s work in ruins, and Hardy would have enjoyed it more than anybody.’


Missing data can wreak havoc with orthodox methods because this changes the sample space, and thus changes not only the sampling distribution of the estimator, but even its analytical form; one must go back to the beginning for each such case. But however complicated the change in the sampling distribution, the change in the likelihood function is very simple. Bayesian methods accommodate missing data effortlessly; in all cases we simply include in the likelihood function all the data we have; and Bayes’ theorem automatically returns the new optimal estimator for that data set.


Lusted, Lee (1968), *Introduction to Medical Decision Making*, Charles C. Thomas, Publisher, Springfield Illinois. Chapter 1 gives a concise summary of Bayesian principles, the other Chapters give many useful Bayesian solutions to important medical problems, with computer source codes. Lee Lusted (1923 − 1994) was a classmate and fellow Physics major of the writer, at Cornell College many years ago. Then we followed surprisingly common paths, at first unknown to each other. Lusted went into microwave radar countermeasures at the Harvard Radio Research Laboratory, the writer into radar target identification at the Naval Research Laboratory, Anacostia, D. C. After WWII, Lusted enrolled in the Harvard Medical School for an M. D. degree, the writer in the Princeton University Graduate school for a Ph. D. Degree in Theoretical Physics; we were both interested primarily in the reasoning processes used in those fields. Then we both discovered, independently, Bayesian analysis, saw that it was the solution to our problems (a sane physician is concerned, obviously, not with any ‘ensemble’ of patients, but with a single patient who presents a unique case unlike any other; likewise a sane physicist is not concerned with any ensemble of physical situations, but with a single incompletely known one) and devoted the rest of our lives to it. At essentially the same time, Arnold Zellner (1971) followed a similar course, moving from Physics to Economics. Thus the modern Bayesian influence in three quite different fields arose from physicists, all of nearly the same age and tastes.

Fish, R. C. Summerfelt & G. E. Hall, editors, Iowa State University Press, Ames, Iowa, pp. 371–384. A computer program for deconvolving mixtures of normal and other distributions. The program, ‘MIX 3.0’ is available from: Ichthus Data Systems, 59 Arkell St., Hamilton, Ontario, Canada L8S 1N6. In Chapter 7 we note that the problem is not very well posed; Ichthus acknowledges that it is ‘inherently difficult’ and may not work satisfactorily on the user’s data. See also Titterington, et al (1985).

Machol, R. E., Ladany, S. P. & Morrison, D. G. (eds), (1976), Management Science in Sports, Vol. 4, TIMS Studies in the Management Sciences, North-Holland, Amsterdam. Curious applications of probability theory, leading to even more curious conclusions. See the advice about which points are most important in a tennis match.


Martin, R. D. & D. J. Thompson (1982), “Robust–Resistant Spectrum Estimation”, Proc. IEEE, 70, pp. 1097-1115. Evidently written under the watchful eye of their mentor John Tukey, this continues his practice of inventing a succession of ad hoc devices based on intuition rather than probability theory. It does not even acknowledge the existence of Maximum Entropy or Bayesian methods. To their credit, the authors do give computer analyses of several data sets by their methods – with results that do not look very encouraging to us. It would be interesting to acquire their raw data and analyze them by methods like those of Brethorst (1988) that do make use of probability theory; we think that the results would be vastly different.


Middleton, D. (1960), An Introduction to Statistical Communication Theory, McGraw–Hill Book Co., New York. A massive work (1140 pages) with an incredible amount of mathematical material. The title is misleading, since the material really applies to statistical inference in general. Unfortunately, most of the work was done a little too early, so the outlook is that of sampling theory and Neyman–Pearson decision rules, now made obsolete by the Wald decision theory and Bayesian advances. Nevertheless, the mathematical problems – such as methods for solving singular integral equations – are independent of one’s philosophy of inference, so it has much useful material applicable in our current problems. One should browse through it, and take note of what is available here.


Contains penetrating historical remarks about the relation of Laplace and Boole, noting that those who have quoted Boole in support of their attacks on Laplace, may have misread Boole's intentions.


Moore, G. T. & Scully, M. O. eds (1986), *Frontiers of Nonequilibrium Statistical Physics*, Plenum Press, N. Y. Here several speakers affirmed their belief, on the basis of the Bell inequality experiments, that “atoms are not real” while maintaining the belief that probabilities are objectively real! We consider this a flagrant example of the Mind Projection Fallacy, carried to absurdity.


Munk, W. H. & Snodgrass, F. E. (1957), “Measurements of Southern Swell at Guadalupe Island”, Deep–Sea Research, 4, pp 272–286. This is the work which Tukey (1984) held up as the greatest example of his kind of spectral analysis, which could never have been accomplished by other methods; to which in turn Jaynes (1987) replied with Chirp Analysis.


Neyman, Jerzy & Pearson, E. S. (1967), *Joint Statistical Papers*, Cambridge Univ. Press. Reprints of the several Neyman–Pearson papers of the 1930’s, originally scattered over several different journals.


_______ (1952), *Lectures and Conferences on Mathematical Statistics and Probability*, Graduate School, U. S. Dept. of Agriculture. Contains an incredible comparison of Bayesian interval estimation vs. confidence intervals. A good homework problem is to locate the error in his reasoning.


of rational inference, and signifies only that the problem was improperly formulated. That is, if you are able to decide that any observation is an outlier from the model that you specified, then that model does not properly capture your prior information about the mechanisms that are generating the data. In principle, the remedy is not to reject any observation, but to define a more realistic model (as we note in our discussion of Robustness). However, we concede that if the strictly correct procedure assigns a very low weight to the suspicious datum, its straight-out surgical removal from the data set may be a reasonable approximation, very easy to do.

Ore, O (1953), Cardano, the Gambling Scholar, Princeton Univ. Press.


Pearson, K. (1892), The Grammar of Science, Walter Scott, London. Reprinted 1900, 1911 by A. & C. Black, London and in 1937 by Everyman Press. An exposition of the principles of scientific reasoning; notably chiefly because Harold Jeffreys was much influenced by it and thought highly of it. This did not prevent him from pointing out that Karl Pearson was far from applying his own principles in his later scientific efforts. For biographical material on Karl Pearson (1857–1936) see Haldane (1957).


——— (1914–1930) The Life, Letters and Labours of Francis Galton, 3 Vols., Cambridge University Press. Francis Galton had inherited a modest fortune, and on his death in 1911 he endowed the Chair of Eugenics at University College, London. Karl Pearson was its first occupant; this enabled him to give up the teaching of applied mathematics to engineers and physicists, and concentrate on biology and statistics.


——— (1921–33), The History of Statistics in the 17th and 18th Centuries, Lectures given at University College, London (E. S. Pearson, editor); Griffin, London (1978).

Penfield, Wilder (1958), Proc. Nat. Acad Sciences (USA), 44, p. 59. Accounts of observations made during brain surgery, in which electrical stimulation of a specific spots on the brain caused the conscious patient to recall various long-forgotten experiences. This undoubtedly true phenomenon is closely related to the theory of the \( A_p \) distribution in Chapter 18. But now others have moved into this field, with charges that psychiatrists are causing their patients – particularly young children – to recall things that never happened, with catastrophic legal consequences. The problem of recognizing valid and invalid recollections seems headed for a period of controversy.


Pfeiffer, R. H. (1948), Introduction to the Old Testament, Harper & Row Publishers, New York. Such a massive work of scholarship concerning what is now known about the writing of the Old Testament that it is hard to imagine that anyone could ever have read it all. But the material is very well organized, so one can quickly locate any particular topic.


Poincaré, H. (1899), “L’Oeuvre Mathématique de Weierstraß”, Acta Math. 22, 1–18. Contains an authoritative account of the relation between the works of Kronecker and Weierstraß, pointing out that the difference was more in taste than in substance; to be contrasted with that of E. T. Bell (1937), who tries to make them mortal enemies.

(1904), Science et Hypothesis, English translation, Dover Publications, Inc., (1952). Poincaré had the gift of being able to say more in a sentence than most writers can in a page. Full of quotable remarks, as true and important today as when they were written.

(1909), Science et Méthode, English translation, Dover Publications, Inc., (1952). Like Kline (1980), a ringing indictment of the contemporary work in mathematics and logic, for which the Bourbakists have never forgiven him. However, in knowledge and judgment Poincaré was far ahead of his modern critics, because he was better connected to the real world.

(1912), Calcul des probabilités, 2nd. edition, Gauthier-Villars, Paris. Contains the first example of the assignment of a probability distribution by the principle of group invariance.

Poisson, S. D. (1837), Recherches sur la Probabilité des Jugements. First appearance of the Poisson distribution.

Pólya, G. (1920), “Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem,” Math. Zeit., 8, 171–181; reprinted in Pólya (1984), Vol. IV. First appearance of the term “Central Limit Theorem” in print. He does not actually prove the theorem (which he attributes to Laplace), but points out a theorem on uniform convergence of a sequence of monotonic functions which can be used to shorten various proofs of it.

(1921), “Über eine Aufgabe der Wahrscheinlichkeitsrechnung betreffend die Irrfahrt im Strassennetz,”, Math. Ann., 84, 149–160. It is sometimes stated that this was the first appearance of the term “random walk”. However, we may point to Rayleigh (1919) and Pearson (1905),


(1984), Collected Papers, 4 Vols. Gian-Carlo Rota, editor, MIT Press, Cambridge MA. Volume IV contains papers on probability theory and combinatorics, several short articles on plausible reasoning, and a bibliography of 248 papers by him. George Pólya always claimed that his main interest was in the mental processes for solving particular problems rather than in generalizations. Nevertheless, some of his results launched new branches of mathematics through their generalizations by others. The present work was influenced by Pólya in more ways than noted in our Preface: most of our exposition is aimed, not at expounding generalities for their own sake, but in learning how to solve specific problems - albeit by general methods.

(1987), The Pólya Picture Album: Encounters of a Mathematician, G. L. Alexanderson, editor, Birkhäuser, Boston. Over his lifetime, George Pólya collected a large picture album with photographs of famous mathematicians he had known, which he took delight in showing to visitors. After his death, the collection was published in this charming book, which contains about 130 photographs with commentary by Pólya, plus a biography of Pólya by the editor.


(1958), The Logic of Scientific Discovery, Hutchinson & Co., London. Denies the possibility of induction, on the grounds that the prior probability of every scientific theory is zero.
Karl Popper is famous mostly through making a career out of the doctrine that theories may not be proved true, only false; hence the merit of a theory lies in its falsifiability. There is an evident grain of truth here, expressed by the syllogisms of Chapter 1; and Albert Einstein also noted this in his famous remark: “No amount of experiments can ever prove me right; a single experiment may at any time prove me wrong.” Nevertheless, the doctrine is true only of theories which assert the existence of unobservable causes or mechanisms; any theory which asserts observable facts is a counter-example to it.


——— (1974), “Replies to my Critics”, in the Philosophy of Karl Popper, P. A. Schilpp, ed., Open Court Publishers, La Salle. Presumably an authoritative statement of Popper’s position, since it is some years later than his best known works, and seeks to address points of criticism directly.

Popper, K. & Miller, D. W. (1983), “A proof of the impossibility of inductive probability”, Nature, 302, 687–88. They arrive at this conclusion by a process that we examined in Chapter 5; asserting an intuitive ad hoc principle not contained in probability theory. Written for scientists, this is like trying to prove the impossibility of heavier-than-air flight to an assembly of professional airline pilots.

Popov, V. N. (1987), Functional Integrals and Collective Excitations, Cambridge Univ. Press. Sketches applications to superfluidity, superconductivity, plasma dynamics, superradiation, and phase transitions. A useful start on understanding of these phenomena, but still lacking any coherent theoretical basis – which we think is supplied only by the Principle of Maximum Entropy as a method of reasoning.


Prenzel, H. V. (1975), Dynamic Trendline Charting: How to Spot the Big Stock Moves and Avoid False Signals, Prentice–Hall, Englewood Cliffs, N. J. Contains not a trace of probability theory or any other mathematics; merely plot the monthly ranges of stock prices, draw a few straight lines on the graph, and their intersections tell you what to do and when to do it. At least, this system does enable one to see the four year Presidential Election cycle, very clearly.


Preston, C. J. (1974), Gibbs States on Countable Sets, Cambridge Univ. Press. Here we have the damnable practice of using the word state to denote a probability distribution. One cannot conceive of a more destructively false and misleading terminology.


Quetelet, L. A. (1835), Essai de Physique sociale.


Raiffa, H. A. & Schlaifer, R. S. (1961), Applied Statistical Decision Theory, Graduate School of Business Administration, Harvard University.

Ramsey, F. P. (1931), *The Foundations of Mathematics and Other Logical Essays*, Routledge and Kegan Paul, London. Frank Ramsey was First Wrangler in Mathematics at Cambridge University in 1925, then became a Fellow of Kings College where among other activities he collaborated with John Maynard Keynes on economic theory. He would undoubtedly have become the most influential Bayesian of the Twentieth Century, but for the fact that he died in 1930 at the age of 26. In these essays one can see the beginnings of something very much like our exposition of probability theory.


Reid, Constance (1982), *Neyman – From Life*, Springer–Verlag, N. Y.


We think it has a bright future, but are not yet prepared to predict just what it will be.


Rothman, Tony (1989), *Science à la Mode*, Princeton University Press. Accounts of what happens when scientists lose their objectivity and jump on bandwagons. We would stress that they not only make themselves ridiculous, they do a disservice to science by promoting sensational but nonproductive ideas. For example, we think that it will be realized eventually that the ‘Chaos’
bandwagon has put a stop to the orderly development of a half–dozen different fields without enabling any new predictive ability. Because, whenever chaos exists, it is surely predicted by the Hamiltonian equations of motion – just what we have been using in statistical mechanics for a Century. The chaos enthusiasts cannot make any better predictions than does present statistical mechanics, because we never have the accurate knowledge of initial conditions that would require. It has always been recognized, since the time of Maxwell and Gibbs, that if we had exact knowledge of a microstate, that would enable us in principle to predict details of future ‘thermal fluctuations’ at present impossible; given such information, if chaos is present, its details would be predicted just as well. But in present statistical mechanics, lacking this information, we can predict only an average over all possible chaotic behaviors consistent with the information we have; and that is just the traditional thermodynamics.


_______ (1981), The Writings of Leonard Jimmie Savage – A Memorial Selection, Published by the American Association of Statistics and the Institute of Mathematical Statistics. Jimmie Savage died suddenly and unexpectedly in 1971, and his colleagues performed an important service by putting together this collection of his writings that were scattered in many obscure places and hard to locate. Some personal reminiscences about him are in Jaynes (1984b) and Jaynes (1985e).


Schnell, E. E. (1960), “Samuel Pepys, Isaac Newton and probability”, Am. Stat. 14, 27–30. From this we learn that both Pascal and Newton had the experience of giving a correct solution and not being believed; the problem is not unique to modern Bayesians.


——— (1947), “The Foundation of the Theory of Probability”, Proc. Roy. Irish Acad. (A), pp 51 – 66; 141 – 146. Valuable today because it enables us to add one more illustrious name to the list of those who think as we do. Here Schrödinger declares the “frequentist” view of probability inadequate for the needs of science and seeks to justify the view of probability as applying to individual cases rather than ‘ensembles’ of cases, by efforts somewhat in the spirit of our Chapters 1 and 2. He gives some ingenious arguments but, unknown to him, these ideas had already advanced far beyond the level of his work. He was unaware of Cox’s theorems and, like most scientists of that time with Continental training, he had apparently never heard of Thomas Bayes or Harold Jeffreys. He gives no useful applications and obtains no theoretical results beyond what had been published by Jeffreys eight years earlier. Nevertheless, his thinking was aimed in the right direction on this and other controversial issues.

——— (1948), Statistical Thermodynamics, Cambridge Univ. Press.

Schuster, A. (1897), “On Lunar and Solar Periodicities of Earthquakes”, Proc. Roy. Soc. 61, pp. 455–465. This marks the invention of the periodogram and could almost be called the origin of orthodox significance tests. He undertakes to refute some claims of periodicities in earthquakes, by considering only the sampling distribution for the periodogram under the hypothesis that no periodicity exists! He never considers: what is the probability of getting the observed data if a periodicity of a certain frequency does exist? Orthodoxy has been following this nonsensical procedure ever since. We show here that evidence for periodicity is contained in the shape of the periodogram, not its sampling distribution. But to show this requires the elimination of nuisance parameters in a way that orthodox ideology cannot comprehend.


Shafer, G. (1982), “Lindley’s Paradox”, J. Am. Stat. Assn., 77, 325–334. Apparently, Shafer was unaware that this was all in Jeffreys (1939; p. 194) some twenty years before Lindley. But Shafer’s other work had made it clear already that he had never read and understood Jeffreys.


Shewhart, W. A. (1931), Economic Control of Quality of Manufactured Products, van Nostrand, New York.


Shore, J. E. and Johnson, R. W. (1980), “Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy”, IEEE Trans. Information Theory IT-26, 26–37. Many different choices of axioms all lead to the same actual algorithm for solution of problems. The authors present a different basis from the one first proposed (Jaynes, 1957), But we stress that maximum entropy and minimum cross-entropy are not different principles; a change of variables converts one into the other.

Siegmann, D., (1985) Sequential Analysis, Springer. No mention of Bayes' theorem or optional stopping!!


Smith, W. B. (1905), “Meaning of the Epithet Nazorean”, The Monist, 15, 25–95. Concludes that prior to the Council of Nicea ‘Nazareth’ was not the name of a geographical place; it had some other meaning.


Spinoza, B., Ethics, part 2, Prop. XLIV: “De natura Rationis no est res, ut contingentes; sed, ut necessarias, contemplari.”

Spitzer, F. (1964), Principles of Random Walk, van Nostrand, N. Y. Background history and present status.


——— (1964), “Inadmissibility of the usual estimate for the variance of a normal distribution with unknown mean”, Ann. Inst. Stat. Math., 16, 155–160. Stein’s inadmissibility discoveries, while shocking to statisticians with conventional training, are not in the least disconcerting to Bayesians. They only illustrate what was already clear to us: that the criterion of admissibility, which ignores all prior information, is potentially dangerous in real problems. Here that criterion can reject as ‘inadmissible’ what is in fact the optimal estimator.


——— (1986), The History of Statistics, Harvard Univ. Press. A massive work of careful scholarship, required reading for all students of the subject. Gives full discussions of many topics that we touch on only briefly.


Tax, S., Editor (1960), Evolution After Darwin, 3 Vols., University of Chicago Press. Volume 1: The Evolution of Life; Volume 2: The Evolution of Man; Volume 3: Issues in Evolution. A collection of articles and panel discussions by many workers in the field, summarizing the state of knowledge and current research directions 100 years after the original publication of Darwin.


Tikhonov, A. N. & Arsenin, V. Y. (1977), \textit{Solutions of Ill-posed Problems}, Halsted Press, New York. A collection of \textit{ad hoc} mathematical recipes, in which the authors try persistently to invert operators which have no inverses. Never perceives that these are problems of \textit{inference}, not \textit{inversion}.

Titchmarsh, E. C. (1937), \textit{Introduction to the Theory of Fourier Integrals}, Clarendon Press, Oxford U. K. The ‘state of the art’ in Fourier analysis just before the appearance of Lighthill (1957), which made all the lengthy convergence theory nearly irrelevant. However, only a part of this classic work is thereby made obsolete; the material on Hilbert transforms, Hermite and Bessel functions, and Wiener–Hopf integral equations, remains essential for applied mathematics.

Titchmarsh, E. C. (1939), \textit{The Theory of Functions}, 2nd edition, Oxford University Press. In Chapter XI the reader may see – possibly for the first time – some actual examples of nondifferentiable functions. We discuss this briefly in Appendix B.


Todhunter, Isaac (1873), \textit{A History of the Mathematical Theories of Attraction and the Figure of the Earth}, 2 vols, Macmillan, London; reprinted 1962 by Dover Press, New York.


Truesdell, C. (1987), \textit{Great Scientists of Old as Heretics in “The Scientific Method”}, University Press of Virginia. The historical record shows that some of the greatest advances in mathematical physics were made with little or no basis in experiment, in seeming defiance of the ‘scientific method’ as usually proclaimed. This just shows the overwhelming importance of creative hypothesis formulation as primary to inference from given hypotheses. Unfortunately, while today we have a well developed and highly successful theory of inference, we have no formal theory at all on optimal hypothesis formulation, and very few successful recent examples of it. A vast amount of fundamental investigation remains to be done here.


\textup{\texteab{同一}} (1977), \textit{Exploratory Data Analysis}, Addison-Wesley, Reading MA. Introduces the word ‘resistant’ as a data-oriented version of ‘robust’.

a sneaky way of committing indecent methodological sins “while modestly concealed behind a formal apparatus.”

(1984), “Styles of Spectrum Analysis”; Scripps Institution of Oceanography Reference Series 84–85, March 1984; pp. 100–103. A polemical attack on all theoretical principles, including Autoregressive models, Maximum Entropy, and Bayesian methods. The “protagonist of maximum entropy” who appears on p. 103 is none other than E. T. Jaynes; further comments on this are in Chapter 22.


Tukey, J. W. & Brillinger, D. (1982), “Spectrum Estimation and System Identification Relying on a Fourier Transform”, unpublished. This rare work was written as an invited paper for the IEEE Special Issue of September 1982 on Spectrum Analysis, but its length (112 pages in an incomplete version) prevented its appearing there. We hope that it will find publication elsewhere, because it is an important historical document. Tukey (1984) contains parts of it.

Twain, Mark (1900), The Complete Short Stories and Famous Essays, P. F. Collier & Son, New York; p. 187. Coming across a French translation of his story: “The Jumping Frog of Calaveras County”, he rendered it back literally into English; the result was hilariously funny, telling a grossly distorted story with the meaning altered at many places.


Valavanis, S. (1959), Econometrics, McGraw–Hill, N. Y. Modern students will find this useful as a documented record of what Econometrics was like under orthodox statistical teaching. The demand for unbiased estimators at all costs can lead him to throw away practically all the information in the data; he just does not think in terms of information content.

Valery–Radot, René (1923), The Life of Pasteur, Doubleday, Page & Co., Garden City, N. Y.


van Dantzig, D. (1957) “Statistical Priesthood (Savage on personal probabilities)”, Statistica Neerlandica 2, 1–16. Younger readers who find it difficult to understand today how Bayesians could have had to fight for their viewpoint, should read this attack on the work of Jimmie Savage. But one should realize that van Dantzig was hardly alone here; his views were the ones most commonly expressed by statisticians in the 1950's and 1960's.


Venn, John (1866), The Logic of Chance, MacMillan & Co., London. Later editions, 1876, 1888. Picks up where Cournot and Ellis left off in the anti–Laplace cause. Some details are given in Jaynes (1986b).


Vorzimmer, P. J. (1970), *Charles Darwin: The Years of Controversy*, Temple Univ. Press, Philadelphia. An account of the evolution of Darwin’s *Origin of Species* from its first publication in 1859 to the sixth and last revision in 1872, as Darwin sought to answer his critics. Some see in this the beginning of the process by which the “naturalist” was replaced by the trained biologist. But others, more perceptive, note that Darwin’s first edition makes a better case for his theory than does the last. The contemporary criticisms that he was answering seem today to be mostly incompetent and not deserving of any reply. Even today, as Stephen J. Gould has noted, attacks on Darwin’s theory only document the author’s misunderstanding of what the theory is. See also Dawkins (1987).

Wald, A. (1941), *Notes on the Theory of Statistical Estimation and of Testing Hypotheses*, Mimeographed, Columbia University. At this time, Wald was assuring his students that Bayesian methods were entirely erroneous and incapable of dealing with the problems of inference. Nine years later, his own research had led him to the opposite opinion.


——— (1950), *Statistical Decision Functions*, Wiley, New York. Wald’s final work, in which he now recognized the fundamental role of Bayesian methods and called his optimal methods “Bayes strategies”.


Walley, P. (1991), *Statistical Reasoning with Imprecise Probabilities*, Chapman & Hall, London. Worried about improper priors, he introduces the notion of a ‘Near-Ignorance Class’ (NIC) of priors. Since then, attempts to define precisely the NIC of usable priors have occupied many authors. We propose to cut all this short by noting that any prior which leads to a proper posterior distribution is usable and potentially useful. Obviously, whether a given improper prior does or does not accomplish this is determined not by any property of the prior alone, but by the joint behavior of the prior and the likelihood function; that is, by the prior, the model, and the data. Need any more be said?

Watson, James D. (1968), *The Double Helix*, Signet Books, New York. The famous account of the events leading to discovery of the DNA structure. It became a best seller because it inspired hysterically favorable reviews by persons without any knowledge of science, who were delighted by the suggestion that scientists in their ivory towers have motives just as disreputable as theirs. This was not the view of scientists on the scene with technical knowledge of the facts, one of whom said privately to the present writer: “The person who emerges looking worst of all is Watson himself.” But that is ancient history; for us today, the interesting question is: would the discovery have been accelerated appreciably if the principles of Bayesian inference, as applied to X-Ray diffraction data, had been developed and reduced to computer programs in 1950? We suspect that Rosalind Franklin’s first “A-structure” photograph, which looks hopelessly confusing to the eye at first glance, if analyzed by a computer program [like those of Bretthorst (1988) but adapted to this problem], would have pointed at once to a double helix as overwhelmingly the most probable structure (at least, the open spaces which say “helix” were present and could be recognized by the eye after the fact). The problem is, in broad aspects, very much like that of radar target identification. For another version of the DNA story, with some different recollections of the course of events, see Francis Crick (1988).

Whittaker, E. T. & Robinson, G. (1924), *The Calculus of Observations*, Blackie & Son, London. Notable because the fake ‘variable star’ data on p. 349 were used by Bloomfield (1976), who proceeded to make their analysis, with absurd conclusions, the centerpiece of his textbook on spectrum analysis.
Wiener, N. (1948), *Cybernetics*, J. Wiley & Sons, Inc., New York. On p. 109, Norbert Wiener reveals himself as a closet Bayesian, although we know of no work of his that actually uses Bayesian methods. But his conceptual understanding of the real world was in any event too naïve to have succeeded. On p. 46 he gets the effect of tidal forces in the earth–moon system backwards (speeding up the earth, slowing down the moon). The statements about the work of Gibbs on pp. 61–62 are pure inventions; far from introducing or assuming ergodicity, Gibbs did not mention it at all. Today it is clear, from the discovery of strange attractors, chaos, etc., that almost no real system is ergodic, and in any event ergodicity is irrelevant to statistical mechanics because it makes no functional difference in the actual calculations. In perceiving this, Gibbs was here a Century ahead of the understanding of others. Unfortunately, Wiener’s statements about Gibbs were quoted faithfully by other authors such as S. Goldman (1953) and Y. W. Lee (1960), who were in turn quoted by others, thus creating a large and still growing folklore. Wiener did not bother to proof-read this work, and many equations are only vague hints as to the appearance of the correct equation.
Wigner, E. (1967), *Symmetries and Reflections*, Indiana University Press, Bloomington. From the standpoint of probability theory, the most interesting essay reprinted here is #15, “The Probability of the Existence of a Self–Reproducing Unit”. Writing the quantum–mechanical transformation from an initial state with (one living creature + environment) to a final state with (two identical ones + compatible environment), he concludes that the number of equations to be satisfied is greater than the number of unknowns, so the probability of replication is zero. Since the fact is that replication exists, the argument if correct would show only that quantum theory is invalid.


Woodward, P. M. (1953), *Probability and Information Theory, with Applications to Radar*, McGraw-Hill, N. Y. An interesting historical document, which shows prophetic insight into what was about to happen, but unfortunately just misses the small technical details needed to make it work.

Wrinch, D. M. & Jeffreys, H. (1919), Phil. Mag. 38, 715-734. This was Harold Jeffreys’ first publication on probability theory, concerned with modifications of the Rule of Succession. He must have liked either the result or the association, because for the rest of his life he made reference back to this paper on every possible occasion. Dorothy Wrinch was a mathematician born in Argentina, who studied at Cambridge University and later taught at Smith College in the United States and, in the words of Jeffreys, “became a biologist”. Her photograph may be seen in Pólya (1987), p. 85. Two later papers by Wrinch and Jeffreys on the same topic are in the Phil. Mag. 42, 369-396 (1921); 45, 368-374 (1923).


Zellner, A. (1971), *An Introduction to Bayesian Inference in Econometrics*, J. Wiley & Sons, Inc., New York. Second edition (1987); R. E. Krieger Pub. Co., Malabar, Florida. In spite of the word “Econometrics” in the title, this work concerns universal principles and will be highly valuable to all scientists and engineers. It may be regarded as a sequel to Jeffreys (1961), carrying on multivariate problems beyond the stage reached by him. But the notation and style are the same, concentrating on the useful analytical material instead of mathematical irrelevancies. Contains a higher level of understanding of priors for linear regression than could be found in any textbook for more than 20 years thereafter.

Zellner, A. (1984), *Basic Issues in Econometrics*, Univ. Chicago Press. A collection of 17 reprints of recent articles discussing and illustrating important principles of scientific inference. Like the previous reference, this is of value to a far wider audience than one would expect from the title. The problems and examples are stated in the context of economics, but the principles themselves are of universal validity and importance. In our view they are if anything even more important for physics, biology, medicine, and environmental policy than for economics. Be sure to read Chap. 1.4, entitled: “Causality and Econometrics”. The problem of deciding whether a causal influence exists is vital for physics, and one might have expected physicists to have the best analyses of it. Yet Zellner here gives a far more sophisticated treatment than anything in the literature of physics or any other ‘hard’ science. He makes the same points that we stress here with cogent examples showing why prior information is absolutely essential in any judgment of this.

Zubarev, D. N. (1974) *Nonequilibrium Statistical Thermodynamics*, Plenum Publishing Corp., New York. An amazing work; develops virtually all the MAXENT partition functional algorithm as an *ad hoc* device; but then rejects the MAXENT principle which gives the rationale for it and explains why it works! As a result he is willing to use the formalism only for a tiny fraction of the problems which it is capable of solving, and thus loses practically all the real value of the method. A striking demonstration of how useful applications can be paralyzed – even when all the requisite mathematics is at hand – by orthodox conceptualizing about probability.