THE ACCUMULATION OF IMPRECISE WEIGHTS OF EVIDENCE

I.J.Good May 26, 2003 July 8, 2003 July 9, 2003

Abstract

A familiar method for modeling imprecise or partially ordered probabilities is to regard them as interval-valued. It is proposed here that it is better to assume a Gaussian form for the logarithm of the odds. To fix the hyperparameters of the Gaussian curve one could make judgements for the quartiles for example. The same comment applies for weights of evidence. The reason for this proposal is that when the pieces of evidence are statistically independent one has additivity and the addition of Gaussian variables is easy to perform. When the pieces of evidence are dependent, there is a more general additivity, or one might be able to allow for interactions of various orders. Possible applications would be to legal trials and to differential diagnosis in medicine, or even for distinguishing between two hypotheses in general.

***8

The three organizers have shown a lot of initiative in arranging this conference and I would like to thank them for inviting me to give a paper. I apologize for not being able to attend in person.

I will start with some background material especially from my own publications, in particular my paper #1515. This was a brief survey of weights of evidence at the second Valencia meeting on Bayesian statistics twenty years ago.

Introduction.

A probability can be a physical (material) probability or chance on the one hand and epistemic probability on the other. An epistemic probability can be either a logical probability (credibility) or a personal one. A personal probability is usually called subjective so as not to sound too personal. Poisson (1837, p.31) distinguished between physical and logical probability and wrote as if personal differences arise only because different people possess different information. Keynes (1921) emphasized that credibilities are only partially ordered and mentioned earlier references in a footnote on his page 5. Partially ordered probabilities obey some obvious transitive properties as also do utilities. There is one exception: a meal of steak can be perceptively no better than one of chicken and chicken can be perceptively no better than one of lobster and yet steak can be perceptively better than lobster. It then follows logically, though not by direct perception, that steak must be better than chicken. (See #1357, a joint paper with T.N.Tideman.) A similar comment applies to partially ordered probabilities or utilities or weights of evidence.

In my 1950 book, #13, I suggested a "black box" theory of probability and emphasized that subjective probabilities are only partially ordered. I hadn't heard of De Finetti's work at that time: it was drawn to the attention of English-reading people by Jimmie Savage (1954). In #230 I used this theory to derive some axioms for imprecise probabilities. Simultaneously and independently Cedric A.B.Smith (1965) showed the self-consistency of the black box theory by using arguments similar to those by which Savage (1954) arrived at axioms for precise probabilities.

Definitions etc.

A proposition is the meaning of a statement. An event is something that might happen or could be imagined to happen. Symbols like E, H, and G denote propositions, usually (but not necessarily) referring to events and hypotheses. A theory or hypothesis is regarded as (approximately) true if its consequences are. A simple hypothesis is one that assigns probabilities to the members of a set of events. The probability of E given or assuming or conditional on H, if this probability is defined, is denoted by $P(E \mid H)$. If G is also "given" or assumed, then the probability is denoted by $P(E \mid H \& G)$ or $P(E \mid H.G)$ or $P(E \mid HG)$. The probabilities might be precise or imprecise. Other people use other notations. It isn't permissible in my philosophy to condition on propositions of zero probability. (Karl Popper wouldn't agree because he argues that universal hypotheses have zero probabilities but this seems to be a mistake as argued for example in

#191 which is a review of Popper's *Logic of Scientific Discovery*. An amendment to #191 is that $\Pi(1 - p_r)$, as well as Πp_r must be convergent and non-zero.) A colon denotes "provided by" and must be distinguished from the vertical stroke as in the notation W(H : E | G), the weight of evidence *in favor* of H *provided by* E *given* G. It will be necessary to discuss later what is meant by "weight of evidence". Non-Bayesians don't like the concept of the probability of H given E so, if they consistently think in terms of probabilities, they shouldn't be judges or jurors. What can "beyond a reasonable doubt" mean to someone who doesn't believe in degrees of belief? The law doesn't state the threshold for conviction, but in a civil case the law seems to be somewhat more explicit.

A weight of evidence, if ordinary English is to be respected, should depend on probabilities and not on utilities. Before rational actions can be taken, allowance must be made for utilities or quasi-utilities which are substitutes for utilities when it is difficult to estimate actual utilities. Weights of evidence, amounts of information, and "explicativities" can themselves be regarded as quasi-utilities. (For explicativities see, for example, *Good Thinking*, Chap. 23, a republication of #1000.)

Weights of evidence and Bayes factors are primarily for discriminating between pairs of hypotheses. If one of them, say the hypothesis of guilt wins easily, then a further hypothesis, say that the accused has been framed, can be entertained. Then again, several hypotheses can all be considered in pairs and the calculation could be completed as in methods for treating an all-against-all sports or chess competition (e.g. #50).

Goddesses of Justice.

In #1715 I used a picture of Themis, the ancient Greek goddess of law and order, modernized by the artist Anna Davidian (Fig. 1). In classical mythology Themis carries a pair of scales and a "horn of plenty". The sword in Davidian's drawing might not be true to the mythology. This was pointed out to me by Nelson A. Blachman. Davidian's beautiful drawing was like historical fiction, partly historical and partly fictional. An example of historical fiction was the film "Enigma" which had exciting vignettes but was extremely misleading. In the much earlier Egyptian mythology the goddess of justice, called Maat, uses scales in the Underworld for weighing the souls or hearts of the dead (Larousse, 1959, p. 41). She, or something she carried, was placed in one pan of the balance and the heart in the other one. If there was perfect equilibrium Osiris rendered favorable judgment and the deceased would mingle freely with the gods and the spirits of the dead, leading a life of eternal happiness. If the deceased was guilty the heart was devoured by a monster called Amemait and presumably there was then death after death.

In Davidian's drawing there are floppy discs which I assume contain weights of evidence for or against a hypothesis such as the guilt or innocence of a suspect. It is a reasonable representation of modern justice. The ancients had no floppy discs although some children today might not know that.

A brief history of the modern view.

(I here quote from #1715, p. 445.) Hume (1748) groped towards a qualitative notion of weights of evidence when he said:

> "No testimony is sufficient to establish a miracle, unless the testimony be of such a kind, that its falsehood would be more miraculous than the fact it endeavors to establish, and even in that case there is a mutual destruction of arguments, and the superior only gives us an assurance suitable to that degree of force which remains after deducting the inferior."

Laplace (1820, pp. 446-461), when considering the probability of testimony, almost explicitly anticipated the concept of a *Bayes factor which is the ratio of the posterior to the prior odds* (not probability) of a hypothesis. (In #1715, p. 445 I mentioned Poisson and De Morgan for related statements.) The famous philosopher C. S. Pierce (1878) came very close to the best definition of weight of evidence, namely the logarithm of a Bayes factor. His remark about weight of evidence, in his rather obscure article, was somewhat of a throw-away line because he was otherwise against Bayesianism (which he called the conceptual approach to probability). The concept of a Bayes factor is explicit in Wrinch & Jeffreys (1921, p. 387) except that they did not use the expressions 'Bayes factor' and 'odds'. A.M.Turing, in World War

II, called it simply a *factor*, but that name would be too ambiguous today. The expression *Bayes factor* is now well entrenched in statistical writings. Minsky & Selfridge (1961) independently used the expression *weight of evidence* in the same sense as used by Pierce and Good.

The Bayes factor *in favor* of a hypothesis H *provided by* evidence E, *given* G all along, is

$$BF(H : E \mid G) = O(H \mid E \,\&\, G) \div O(H \mid G) = P(E \mid H \,\&\, G) \div P(E \mid {\sim}H \,\&\, G), \qquad (1)$$

where O denotes odds and the tilde denotes negation. The right side is not a likelihood ratio in general but it is when H and not H are simple statistical hypotheses. Otherwise it could be called a *Bayesian Likelihood Ratio*. This theorem follows by four applications of the product axiom of epistemic probability. Perhaps this should be called *the fundamental theorem of the philosophy of science.* It is natural to take the logarithm for defining weight of evidence because this gives rise to additivity (appropriate for weights) as in:

$$W(H : E_1 \,\&\, E_2) = W(H : E_1) + W(H : E_2 \mid E_1). \qquad (2)$$

Formula (2) extends easily to the case of several events. For a more formal and convincing derivation of the explication of W see #1515, p.251. All the weights of evidence can be made conditional on another proposition G. The additivity simplifies when the events are statistically independent given H and also given ~H. Examples of statistical independence at least approximately in a criminal trial, are motivation, opportunity and ability. For example, if a victim is shot, a good marksman gets more weight of evidence against his innocence, than would a dog, but the sign would be reversed if the victim were bitten (vampires apart). An alibi decreases opportunity. This example is deliberately oversimplified for the sake of brevity.

The expression *weight of evidence* was used in #13 and again in at least forty of my publications, of course with some repetition of. (See, for example #s 1515 and 1828 and the *two* subject indexes of *Good Thinking.*) As mentioned above, the expression was independently used by Pierce and by Minsky & Selfridge. For ordinary English usage, the Oxford English Dictionary quotes T.H.Huxley (1878) as saying "The weight of evidence appears strongly in favor of the claims of Cavendish". (Huxley is famous for his work on Darwinism and for his educational writings on that topic as in Huxley, 1908.)The technical concept of weight of evidence captures this ordinary usage very well indeed. The technical concept, but not the name, was used by A.M.Turing in World War II for a procedure called *Banburismus* because the stationery used was printed in the town of Banbury. This was one important cryptanalytic procedure for regularly breaking the Naval Enigma. This was an enciphering machine used by the Germans. For more on this topic see #2117H. The unit of weight of evidence was called by Turing a *ban* when the base of the logarithms was 10. This name was analogous to Tukey's name 'bit' in Shannon's theory of communication or information. Turing called one tenth of a ban a *deciban* (db) by analogy with the name *decibel* in acoustics. The *deciban*, or the half-deciban (hdb) is about the smallest discernible weight of evidence. When the base of logarithms is e, Turing called the unit a *natural ban* which is convenient for theoretical purposes.

Relationships between expected weight of evidence and entropy in Shannon's sense are discussed in #1505. One name for expected weight of evidence is *cross-entropy.*

One application of the concept of weight of evidence is to the topics of *necessitude* and *sufficitude.* These are measures of the degrees to which one event F is necessary or sufficient for another event E. They are defined by taking strict necessity and sufficiency and replacing strict implication by weight of evidence. For example, the sufficitude is equal to the weight of evidence against F if E does not occur, given the state of the universe just before F or ~F occurred. The difficulty in pinning down causation occurs because one can never be sure of the state of the universe. For a detailed discussion, with legalistic interpretations, see #2200.

When several small weights of evidence are added together the sum would have roughly a normal distribution. Turing showed in 1940 or 1941 that, when W has a normal distribution, the variance of the sum in natural bans is twice the expectation. This was noticed also by Peterson, Birdsall, & Fox, (1954). When decibans are used this surprising theorem can be expressed in the convenient form that the standard deviation is close to three times the square root of the expectation. If you look at numerical examples this is perhaps the most terrifying theorem in mathematics because it shows how easily evidence can point in the wrong direction. For the case where the weight of evidence is only approximately normally distributed see #221 which deals with false-alarm probabilities. A false alarm could lead to a war.

How Bayesian should a legal trial be?

The use of Bayes's theorem in legal proceedings is still a controversial issue (Hutton, 2003), partly because people are by no means perfect Bayesians. But even dogs are fairly good Bayesians otherwise they wouldn't survive as long as they do. I wonder whether dogs could be used as adjuncts to lie-detectors because of their excellent sense of smell. In elementary education people could be taught the concepts of Bayes factors and weights of evidence. Incidentally it would make them more interested in logarithms.

Let us consider a concrete legal example (#s 2190C, 2230, 2240, 2240A). Alan Dershowitz had argued that wife-battery should be regarded as inadmissible evidence on the grounds that wife-batterers seldom murder their wives. There are degrees of battery. I defined a "standard batterer" as one who batters his wife about once per year. Dershowitz overlooked that the fact that the wife had been murdered by somebody is an extremely important additional piece of evidence. Indeed, by using the hard statistics quoted by Dershowitz himself, the Bayes factor method leads to the conclusion that the husband's odds of guilt are about 10 (or 10 to 1 on) if he is a "standard batterer". Of course this is before other evidence is taken into account. In the O.J.Simpson case there was a lot of other evidence. The main counter-evidence was from the glove that appeared not to fit. But a clipping of film in a Charles Grodin TV performance showed Simpson pulling off the glove with no difficulty at all immediately after having hood-winked the jury.

Of course in general it isn't easy for a juror to estimate initial (prior) odds. It is less difficult to estimate the odds at some intermediate stage of the trial or, for this purpose, by taking some subset of the evidence, not necessarily considering the evidence in the order in which it is presented in court. Then the Bayes factor has to be judged by the rest of the evidence. It is a responsibility of the prosecuting attorney to present the evidence in a fair and appropriate order to alleviate the task of the jury in this two-stage process.

Typically most of the weights of evidence in a legal trial are very imprecise. There might also be some fairly precise weights of evidence evaluated by professional statisticians.

The fact that the accused is in the dock should certainly not be taken as a basis for a judgment of the initial odds. The evidence presented in the trial would overlap with the reasons why the accused was in the dock. To hold it against the accused that he is suspected would be to use the same evidence twice. Even that wouldn't be as wicked as a trial by ordeal as in the notorious Spanish Inquisition --- wickedness perpetrated in the name of God.

Imprecise Weights of Evidence and their accumulation.

A theory of the imprecise, a qualitative theory, should be a generalization of a quantitative theory, a theory of the precise. Qualitative and quantitative theories shed light on each other.

In a familiar model for partially-ordered probabilities the probabilities are interval-valued. In other words there are lower and upper probabilities. But these end-points of intervals, apart from being imprecise or vague, surely have lower probability densities than say the mid-points of the intervals. (These probability densities refer to probabilities of probabilities and might be called probabilities of level two. I now prefer not to call them probabilities of type two because, in my terminology, rationality of type two means rationality in which allowance is made for the cost thinking or calculation.)

Since the ends of the intervals are too arbitrary I prefer a model where imprecise log-odds and weights of evidence have (level-two) normal (Gaussian) distributions. Call this the (level two) *normal* model (for weights of evidence).   This device won't do for probabilities or odds because they don't extend from minus to plus infinity. A normal distribution is fixed by its lower and upper quartiles for example. These have to be judged by the users of the theory. (Whether sextiles are easier to judge is a matter for experimentation.) Non-Bayesian statisticians, who use normal distributions habitually, can hardly complain about the present use. I am merely claiming that this new "normal" model is better than the "interval-valued" model. The new "normal" model has the further advantage that the sums and differences of normal random variable (not *mixtures*) are again normally distributed. To make use of this fact we have to assume that the various weight of evidence are either statistically independent or else we can make each weight of evidence conditional an all those already used. Double use of a single weight of evidence is then impossible because the weight of evidence provided by the second usage would be zero (the Bayes factor would be unity).

Interactions between weights of evidence (#210, Appendix 6), being linear in individual weights, should also satisfy the level-two normal model.

The weights of evidence used by the various members of a jury cannot be confidently combined because distinct people might overestimate or underestimate a given piece of evidence by distinct amounts. This problem occurs in medical diagnosis: see, for example, #755, Section 4. Somebody should eventually do some experiments on this matter.

For this normal model I have relied on the principle that "The real problem in formulating a mathematical model is to find an adequate compromise between realism and mathematical convenience." (#142, p.116.)

Sometimes a tail probability or P-value, or rather its reciprocal, is regarded as a non-Bayesian weight of evidence against a null hypothesis. They can be combined by Fisher's method. It it is careless to ignore the sample sizes as Fisher did in one place. I discussed such matters in #1515 Section 7 and in many other places, but it would make the present paper too long to go into details.

The accumulation of imprecise weights of evidence

I. J. Good July 4, 2003

Abbreviations: AMS- Annals of Mathematical Statistics
Biom- Biometrika ("Biometrics" is spelled out)
BJPS- British Journal for the Philosophy of Science
BR- Book Review
I- Informal
IP- Informal Paper
JASA- J. American Statistical Association
JIP- Joint Informal Paper
JLMS- J. London Mathematical Society
JP- Joint Paper
JRNSS- J. Royal Naval Science Service
JRSS- J. Royal Statistical Society
JSCS- J. Statist. Comput. & Simul.
JSPI- J. Statist. Planning and Inference
MR – Mathematical Reviews
MTAC- Mathematical Tables and other aids to Computation
P- Paper
PCPS- Proc. Cambridge Philosophical Society. Later became the Mathematical Proceedings of the Cambridge Philosophical Society.
QJM- Quarterly J. Mathematics Oxford
TR- Technical Report (Stat. Dept. Va. Tech.)
§- A merit of this paper is its brevity
*- Book

* 13. Probability and the Weighing of Evidence (London, Charles Griffin; New York, Hafners; 1950, pp. 119). (Copyright expired.)

P 142. IJG and K. Caj Doog. "A paradox concerning rate of information", Information and Control 1 (1958), 113-126. (See #'s 192 and 210.)

BR 191. "Review of K.R. Popper, "The logic of scientific discovery", MR 21 (1960), 1171-1173.

P 192. "A paradox concerning rate of information: corrections and additions", Information and Control 2 (1959), 195-197. (See #142.)

P 210. "Effective sampling rates for signal detection: or can the Gaussian model be salvaged?" Information and Control 3 (1960), 116-140. (See #'s 142 and 192.)

P 221. "Weight of evidence, causality, and false-alarm probabilities", Fourth London Symp. on Information Theory, Buttersworths, (1961), pp. 125-136.

JP 755. I.J. Good and W.I. Card, "The diagnostic process with special reference to errors", Methods of Information in Medicine 10, No. 3 (July 1971), 176-188.

P 1000. "Explicativity: a mathematical theory of explanation with statistical applications", Proc. Roy. Soc. (London) A 354 (1977), 303-330. Republished in #1161 and 1364.

Rep.P 1161. Republication of #1000, with three misprints corrected, in *Bayesian Analysis in Econometrics and Statistics: Essays in Honor of Harold Jeffreys* (ed. Arnold Zellner; North Holland: Amsterdam, 1980), 21-34.

* 1364. Good Thinking: The Foundations of Probability and its Applications. Univ. of Minnesota Press (1983, Dec.), pp. 332 + xviii.

BR 1505. Review of "The Maximum Entropy Formalism", Raphael D. Levine and Myron Tribus, eds. (1979).

1505A. A somewhat shortened version of #1505, JASA 78 (Dec. 1983), 987-989.

P 1515. "Weight of evidence: a brief survey", in Bayesian Statistic 2: Proceedings of the Second Valencia International Meeting, September 6/10, 1983. (J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith, eds.; New York: North Holland, 1985), 249-269 (including discussion).

1515B. Errata etc. to #'s 1515 and 1515A (includes some additional discussion of Seidenfeld's contribution to the discussion).

BR 1635. Review of J. Kadane (ed.) Robustness of Bayesian Analyses, Volume 4 of studies in Bayesian Econometrics, ed. by A. Zellner and J.B. Kadane; JASA 80 (Dec. 1985), 1067 to 1068.

P 1660. "Scientific method and statistics", Encyclopedia of Statistical Sciences Vol. 8 (1988), 291-304. (New York: Wiley)

P 1715. "Speculations concerning the future of statistics," in the conference Foundations and Philosophy of Probability and Statistics in honor of I.J. Good, 1987 May (K. Hinkelmann, ed.) in the Special Issue on Foundation of Statistics and Probability, JSPI 25, No. 3 (July 1990), 441-446. [Note: My reply to Barnard's Comment is #1751.]

P 1729. "The interface between statistics and the philosophy of science", for the Eighth International Congress on the Logic, Methodology, and Philosophy of Science VIII (J.E.Fensted, I.T.Frolov, and R.Hilpinen, eds.; Amsterdam North Holland & Elsevier; 1989), invited paper sent by videotape for the Moscow conference held in August 1987.

P 1729A. "The interface between statistics and philosophy of science", Statistical Science 3, (1988), 386-412 (with discussion by Patrick Suppes, George Barnard, James O. Berger and David L. Banks).

R 1761. Review of Dempster, A.P., "Probability, evidence, and judgement", in Bayesian Statistics 2 (eds. J.M. Bernardo; H.H. DeGroot; D.V. Lindley; and A.F.M. Smith; Elsevier Science, 1985), pp. 119-132; MR 88i (Sept.), p. 4839, Rev. #62003.

P 1776. "On the combination of pieces of evidence", C311 in JSCS 31, No. 1 (1989), 54-58.

IP 1784. "Weight of evidence and a compelling metaprinciple", C319 in JSCS 31, No. 2 (1989), 121-123.

P 1828. "Weight of evidence and the Bayesian likelihood ratio", Chapter 3 of The Use of Statistics in Forensic Science (C.G.G. Aitken and D.A. Stoney, eds. Ellis Horwood, 1991). Also as a Tech Rep. 89-16.

1860. "Brief comments on some of I.J. Good's publications on statistics and allied topics" (March 14, 1990), pp. 7

MR 1896. Review of Shafer, Glenn. "The unity and diversity of probability." In Statistical Science 5 (1990), 435-462 (with discussion); MR 92d (April), p. 1774, Rev. #010.

P & TR 2117H. "Enigma and Fish," the version in the paperback edn. of Codebreakers (1994), pp. 149-166. Contains an important correction at the top of page 156 supplied by Joan Clarke

Murray. O.U.P didn't publish the fact that the paperback was better than the hardback because they didn't want to undermine the sale of the hardback which was a best seller.

P 2190C. "The probability that the batterer was the murderer". The editors changed the title to "When a batterer turns murderer", Nature 375 (June 15, 1995), 541. (See #2230.)

P 2200. "Legal responsibilities and causation" (Jan. 1997), in Machine Intelligence 15, Furukawa, K; Michie, D.; and Muggleton, S.; eds. 1999, 25-59. (Oxford University Press.)

2200B. "Causation in the law: necessitude and sufficitude", a lecture at the 75th Annual Meeting of the Virginia Academy of Science. Abstract in Va. J. Sc. 48, No. 2 (summer 97), p. 160.

2200C. Errata for #2200, and offer of a prize.

P 2230. "When batterer becomes murderer" Nature 381 (June 6, 1996), 481. (See #2190C, 2240.)

TR & P 2240. "Bayes factors, batterers, murderers, and barristers", to be translated into Italian for KOS (Milan.) TR96-4. (See #2230.)

TR & P 2240A. Italian translation of #2240. KOS, nuova ser. No. 131/132 (Agosto/Septembre 1996), 25-29. "Fattori di Bayes, Mariti violenti, assassini ed arrocati", TR 96-4.

BR & P 2264. "Bayes factors, weights of evidence and the law", a book review article of Joseph B. Kadane & David A. Schum, "A probabilistic Analysis of the Sacco and Vanzetti Evidence" (New York: Wiley, 1966). JSPI 64 (1997), 171-191.

I 2264AA. Erratum to #2264, J. Statist. Plann. Inference 64 (1997), 171-191. In JSPI 75 (1998), 173. Neither I nor the book review editor were responsible for any of the ten known misprints.

IP 1357. IJG and T.N. Tideman, "The relevance of imaginary alternatives", C93 in JSCS 12 (1981), 313-315.

Hume, David. (1748). *An Inquiry Concerning Human Understanding.*

Hutton, Jane. A report of a meeting on April 10, 2003 held at the Royal Society, London, to discuss the Justice system. It was organized by the Forensic Science Society, the Council for the Registration of Forensic Practioners and the Expeert Witness Institute. *RSS News*, June 2003, page 11.

Huxley, Thomas Henry. *Twelve Lectures and Essays* (Watts and Co., London, 1908).

Keynes, J.M. (1921). *A Treatise on Probability* (Macmillan, London).

Laplace, P.S. de (1812/1820). *Théorie Analytique des Probabilités, 3rd edn.* (Courcier, Paris).

Larousse, *New Larouse Encyclopedia of Mythology* (Hamlyn, London, 1959).

Minsky, M. & Selfridge, O.G. (1961), "Learning in random nets", in *Information Theory: Fourth London Symposium* (Colin Cherry, ed.; Butterworths, London), 335-347.

Peirce, C.S. (1878). "The probability of induction", *Popular Science Monthly*; republished in *The World of Mathematics* 2 (James R. Newman ed.; Simon and Schuster; New York, 1956), 1341-1354.

Peterson, W.W., Birdsall, T.G. & Fox, W.C. (1954). "The theory of signal detectability" *Trans. Inst. Radio Engrs PGIT-4, 171-212.*

Poisson, S.D. (1837). *Recherches sur la probabilités des jugements en matières criminelle et matières civiles , précédées des règles générales du calcul des probabilités* (Bachelier; Paris).

Popper, Karl R. (1959), *The Logic of Scientific Discovery* (Hutchinson, London).

Savage, L.J. (1954). *Foundations of Statistics* (Wiley; New York).

Smith, C.A.B. (1961). "Consistency in statistical inference and decision", *JRSS B* 23, 1-25 (with discussion).

Wrinch , Dorothy & Jeffreys, Harold (1921), "On certain fundamental principles of scientific discovery", *Philos. Mag.*, ser. 6, 42, 369-390.