genius of Keynes was applied elsewhere, and the solitary Edgeworth was at the end of his long career.

(4) Textbooks on probability by the mathematicians Whitworth, Burnside and Coolidge gave examples on the probabilities of causes suitable only for the examination room. In view of the title of *SMSI*, there are better reasons for inspecting how inverse probability was treated in textbooks on statistics, or on topics that are statistical in nature. A short list for the period between 1880 and 1930 might include the following books, detailed references for which are scarcely necessary: M. Merriman (1884); A. L. Bowley (1901); G. U. Yule (1911); D. Brunt (1917); E. T. Whittaker and G. Robinson (1924); R. A. Fisher (1925); H. L. Rietz, (1927). The choice of *Statistical Methods for Research Workers* seems appropriate. This book made a fundamental break with tradition (Yates, 1951), and successive editions tolled the death knell of inverse probability for all to hear. The mathematician Neyman seems to have been rather hard of hearing. But Harold Jeffreys firmly rejected the claim and he carried the banner of Bayes and Laplace until the next generation was ready to take over.

ADDITIONAL REFERENCES

KRUSKAL, W. (1980). The significance of Fisher: A review of *R. A. Fisher: The Life of a Scientist. J. Amer. Statist. Assoc.* **75** 1019–1030.

PEARSON, E. S. (1968). Some early correspondence between W. S. Gosset, R. A. Fisher and Karl Pearson, with notes and comments. *Biometrika* **55** 445–457.

PEARSON, E. S. (1974). Memories of the impact of Fisher's work in the 1920s. *Internat. Statist. Rev.* **42** 5–8.

WELCH, B. L. (1958). "Student" and small sample theory. *J. Amer. Statist. Assoc.* **53** 777–788.

YATES, F. (1951). The influence of *Statistical Methods for Research Workers* on the development of the science of statistics. *J. Amer. Statist. Assoc.* **46** 19–34.

Comment

G. A. Barnard

In drafting these comments I have had the advantage of seeing Robin Plackett's, with which I broadly agree. Matters are indeed complex.

I beg to differ from Zabell when he writes that in 1930 "Fisher and Neyman *simultaneously* (my stress, G. A. B.) administered a nearly lethal blow to Bayesian statistics, one from which it was not to recover until the publication ... of Savage's *Foundations of Statistics* in 1954." Neyman's continued interest in Bayesian methods in 1929, correctly noted by Zabell, is hardly consistent with his having shared in giving them a near lethal blow the following year. But Fisher's rejection of inverse *probability*, in the sense used here, is already quite clear in the paper of 1912 to which Zabell refers. The most important difference between 'probability' and Fisher's 'likelihood' as a measure of credibility of statistical hypotheses is that 'likelihood' does not obey the addition laws—as Fisher was wont to say, "the likelihood of H or H'" is, like "the height of Peter or Paul," meaningless unless it is specified which of the two is meant. In the final paragraph of his 1912 paper, Fisher specifically says that what he has been calling "probability" is not to be understood as capable of summation over a set of alternative hypotheses. True, he does not put forward the term 'likelihood' until 1921, but the difference of concept is already there in 1912.

Fisher clearly persuaded Egon Pearson, who in turn eventually persuaded Neyman to abandon Bayesian methods, though, unlike Pearson *fils*, Neyman never accepted likelihood as a valid measure of credibility distinct from probability. Neyman's view of the Neyman–Pearson theory had strong "decision" aspects, while Pearson's view was always more flexible.

But "eclipse" does not seem appropriate to describe the state of a theory which, through the 1930's and later continued to have the support, not only of Jeffreys, but also of such other leading users of statistics as Haldane and Gini. In 1940 Deming caused to be published a reprint of Bayes' paper of 1763, and in his introduction E. C. Molina makes it clear that Bayes' ideas continued to demand attention. Frank Yates' contribution to discussion of a paper of mine in 1946 shows Fisher's most distinguished co-worker in statistics agreeing with a Bayesian approach to problems of a certain type. When Maurice Frechet organized a discussion on statistical inference for the 1949 Paris International Congress on the History and Philosophy of Science, it was natural for him to invite

G. A. Barnard is Professor Emeritus, University of Essex. His mailing address is: Mill House, 54, Hurst Green, Brightlingsea, Colchester, Essex CO7 0EH, United Kingdom.

Fisher, Neyman *and* de Finetti as principal speakers, though Fisher was unable to attend. I. J. Good's Bayesian *Probability and the Weighing of Evidence* was published in 1950. The decade beginning in 1949 saw the publication in the Royal Statistical Society's Journal of a series of discussions of foundational questions in which various types of Bayesian and other approaches to inference were put forward.

It was in reaction to such discussions and to criticisms of his 'fiducial argument' that Fisher, at the age of 65, was induced by his friends to write *Statistical Methods and Scientific Inference* as a general account of his views up to that time. These continued to develop, as the successive editions show. Those of us who read many of the chapters in draft took Chapter II, as its title indicates, to be a sketch of "The Early Attempts and their Difficulties" *as Fisher saw them.* Viewed in this way, it seems to me to stand up better than Zabell suggests. After Boole discovered his "general theorem of development in Logic" he saw, as did Carnap nearly a century later, that use of his theorem may sometimes enable us to enumerate "equally possible cases" in a nonarbitrary way. But the footnote to his later paper which Zabell quotes makes clear that Boole did not retract his statement in "The Laws of Thought," that "when the defect of data is supplied by hypothesis, the solutions will, in general, vary with the nature of the hypotheses assumed." Section 21 of Chapter XX of Boole's book remains a very powerful caution against the abuse of Bayes' theorem in areas of natural science where objectivity is a primary aim. He indicates very clearly what we can, and what we cannot, derive from the theorem.

The attention Fisher pays to Venn and to Chrystal seems odd nowadays. While Venn's book was popular and influential in its day, his exposition is shallow compared with that of Boole or of the Reverend Leslie Ellis. But Venn was President of Fisher's Cambridge College throughout Fisher's student career, while a mastery of Chrystal's *Algebra* was essential to success in the early part of the Mathematical Tripos. Its "reign" in this respect was only just ending when I sat for the Cambridge scholarship examinations in 1932. This goes some way to account for Fisher's use of the word "dramatic" in the address to a broad audience at Michigan State University of 1957. The misspelling there of "Crystal" suggests that Fisher's proofreading was there somewhat below his usual standard.

As Plackett says, Fisher's work needs to be read with caution. Care is also needed, as is indicated, for example in Barnard (1987) where failure by both D. V. Lindley and myself to notice a cross heading showing that the case in question was one where the available observations were of two *different* kinds led to serious misunderstandings.

Plackett has pointed to the rediscovery of Mendelism in 1900 as an influence on Fisher's thinking. It is to be regretted that Fisher inserts his highly significant reference to Mendelism into his section on Boole, instead of making clear its importance by devoting to it a section on its own. To paraphrase Fisher's example, we are interested in a black mouse, B, which may be homozygous ($\theta = 1$) or heterozygous ($\theta = 2$). To test for heterozygosity, we mate B with a brown mouse and obtain the result E: All seven offspring are black. Mendel's laws tell us that $\Pr\{E \mid \theta = 1\} = 1$ while $\Pr\{E \mid \theta = 2\} = 1/2^7$. If π denotes the prior probability that B is heterozygous, i.e., that $\theta = 2$, we have, by Bayes' theorem,

$$\Pr\{\theta = 2 \mid E\} = \frac{\pi/128}{\pi/128 + (1 - \pi) \cdot 1}.$$

If one of B's parents was known to be brown, or if each of its two black parents was known to have produced some brown offspring, then we would know from Mendel's laws that $\pi = 2/3$, and given the experimental result E we could then infer that the probability that B is heterozygous is $1/65$. As Fisher says, "cogent knowledge a priori would have been available and the method of Bayes could properly be applied. But if knowledge of the origin of the mouse tested were lacking, no experimenter would feel he had warrant for arguing as if he knew that of which in fact he was ignorant." All we can say on the basis of the result E alone is that the likelihood ratio for $\theta = 1$ against $\theta = 2$ is $128/1$. Computations such as these are routine in pedigree analysis. They would have been familiar to young Fisher, who entered Cambridge as an undergraduate the year after Bateson, Mendel's great champion, was appointed to the chair of biology. Young Fisher soon became himself a champion of Mendelian ideas and the fact that "likelihood" does not obey the addition rule of probability would thus have been obvious in this context. If we knew π, we could specify the probability that the result of a further crossing of B with a brown mouse would be brown; but without knowledge of B's provenance this would not be possible. The two available likelihoods of $\theta = 1$ and $\theta = 2$ could not be meaningfully added.

In the 1958 talk (*Collected Works* 272) to which Zabell refers, Fisher stresses that the interest in the theory of probability shown by the "old masters" Fermat, Pascal, et al. was associated not only with the high social status of gamblers but also with technical developments that enabled highly accurate unbiased dice, etc. to replace primitive knucklebones. A throw of such unbiased dice provides a model of an "experiment" in which there are several possible outcomes, each arising with a precisely defined probability. Mendelism was the first theory arising in natural science for which such experiments form the standard model—given the genetic constitutions of the animals or plants involved in a cross, the probabilities that a given offspring will have a given genotype are precisely specified. Later in this century, we saw the development of quantum mechanics in which probabilistic predictions of experimental results form the very basis of the theory. So revolutionary was this idea that Einstein, to the end of his life, refused to accept that the quantum-theoretical description of physical reality could be complete; yet the evidence is very strong that it is so. Mendelism and the quantum theory have provided an intellectual climate in which the notion of a repeatable experiment having a number of alternative outcomes, each with precisely specified "experimental probability" is now commonplace. Our formalizations of probability theory have not kept pace with these developments, so that, for instance, there is as yet no agreed notation which serves to distinguish the meanings of the | in $\Pr\{E \mid H\} = p$ and $\Pr\{E \mid F\} = q$, where H is a statistical theory from which it is *deduced* that the probability of E is p (as in the example above when we deduce that $\pi = 2/3$), and where F denotes an incomplete specification of an experimental result and E provides a more complete specification of the same result. In the second case, we can define $\Pr\{E \mid F\}$ as $= \Pr\{E \ \& \ F\}/\Pr\{F\}$, but such a definition makes no sense in the first case.

To understand Fisher's thinking on probability, it is necessary to understand that he always thought of himself as primarily a natural scientist, concerned to explore and test the consequences especially of theories such as Mendel's. For Fisher "probability" was, by definition, "experimental probability." The use of this concept in genetics and in quantum physics provided a philosophical ground in which the idea of a "statistical model" could flourish. The dominant position Fisherian statistics achieved between 1925 and 1965 is surely due to the fact that *Statistical Methods for Research Workers* was directed at those who wished to use such models, and this period saw the successful application on a wide scale of the same type of model in agriculture, biometrics and industrial research and production.

In suggesting that 1954 saw the beginning of a "return" to inverse probability, it seems to me that Zabell is underestimating the novelty of the ideas of Ramsey and de Finetti, whose mathematical theory of "personal probability" is something of which there is no counterpart in 19th century thought. Equally, Jeffreys' theory of "probability" has no earlier counterpart, in that it makes a nearly successful attempt to formalize the way in which, in natural science, the notion of "simplicity" operates in the formulation of scientific laws. For example, Jeffreys' theory provides reasons why the Mendelian ratios (in the absence of differential viability, etc.) should be taken as rational numbers like 2/3 rather than "approximately 0.67" or such other real numbers. Again, the "decision" aspects of Neyman's approach, leading to Wald's general formulation of decision theory, are new in this century apart from the hints given by Gauss in a special context. On the other hand the "logical probability" concept of Keynes and Carnap does have its counterpart in the 19th century, in Laplace and in Boole and the ideas criticized by, for example, Bertrand.

We now have before us the exciting tasks of exploring the interconnections of all these ideas and many more. For example, in a clinical trial, the "decision" to terminate the trial at some stage has decision-theoretical aspects that make a Bayesian analysis helpful; but given that the trial has been terminated, the data represent a contribution to medical science that may better be looked at from a Fisherian point of view. Such data often need to be combined with data from other trials in which the relevant parameter values may be expected to differ in magnitude though not in sign, giving rise to problems of a type intermediate between those we call "significance tests" and those we call problems of estimation. And a clinician who has to choose a treatment for a given patient, taking trial results into account along with other information specific to the patient, can be helped by sophisticated Bayesian analyses such as those being developed by Spiegelhalter and others. In industrial research, experiments may naturally group themselves into sets—for example, of experiments to estimate the rate constants of chemical reactions to be used in industrial processes. For such a set, the collection of rate constants may be regarded as a population having a distribution that can be regarded as approximately known; such a distribution can be usefully introduced as an approximate prior into the analysis of each experiment. It will ill become us to allow our continuing and developing attempts to come to terms with the Protean concepts of uncertainty to become bogged down in battles of an Athanasian versus Arian type.

ADDITIONAL REFERENCES

BARNARD, G. A. (1987). R. A. Fisher—a true Bayesian? *Internat. Statist. Rev.* **55** 183–189.

SPIEGELHALTER, D. J. (1986). Probabilistic reasoning in predictive expert systems. In *Uncertainty in Artificial Intelligence* (L. N. Kanal and J. F. Lemmer, eds.) 47–68. North-Holland, Amsterdam.

YATES, F. (1946). Discussion of "Sequential tests in industrial statistics," by G. A. Barnard. *J. Roy. Statist. Soc. Supp.* **8** 24.