restricted to cases in which the notion of 'chance' is involved, its domain is narrower than that of the subjective Bayesians; at the same time it is more explicit in its application to problems of inference in the natural sciences.

The last chapter shows signs of having been written some time after the earlier ones, and it seems to shift the emphasis in places. For example, on page 222 Hacking comes near to discussing a 'goodness of fit' situation, and says 'My theory of statistical support does not attempt rigorous analysis of the reasoning here. . . . The theory of statistical support cannot judge the force with which an experiment counts against a simplifying assumption.' To this extent he appears to agree with the comment made above on his treatment of tests of significance. Again, on page 219, Hacking appears to entertain the possibility of something corresponding to the idea of 'prior likelihood' of 'acceptability' referred to above, and he explicitly refers to the point made by Fraser (and also by the present writer) that group invariance and other structural features of an experimental set-up may be relevant to its statistical interpretation.

It is clear that in this area there is much further exploration to be done. Hacking's book remains an invaluable guide book for anyone willing to join in this task.

G. A. BARNARD
University of Essex

REFERENCES

ANSCOMBE, F. J. [1962]: 'Tests of Goodness of Fit', *Journal of the Royal Statistical Society (B)*, **25**, pp. 81–94.

BARNARD, G. A. [1947]: Review of Abraham Wald: 'Sequential Analysis', *Journal of the American Statistical Association*, **42**, pp. 658–64.

BARNARD, G. A. [1949]: 'Statistical Inference', *Journal of the Royal Statistical Society (B)*, **11**, pp. 115–49.

BARNARD, G. A. [1950]: 'On the Fisher–Behrens Test', *Biometrika*, **37**, pp. 203–7.

BARNARD, G. A. [1951]: 'The Theory of Information', *Journal of the Royal Statistical Society (B)*, **13**, pp. 46–64.

BARNARD, G. A., JENKINS, G. M. and WINSTEN, C. B. [1962]: 'Likelihood Inference and Time Series', *Journal of the Royal Statistical Society (A)*, **125**, pp. 321–72.

FISHER, R. A. [1925*a*]: 'Theory of Statistical Estimation', *Proceedings of the Cambridge Philosophical Society*, **22**, pp. 700–25.

FISHER, R. A. [1925*b*]: *Statistical Methods for Research Workers.*

RÉNYI, A. [1955]: 'On a New Axiomatic Theory of Probability', *Acta Mathematica Academiae Scientiarum Hungaricae*, **6**, pp. 285–335.

ROBBINS, H. [1952]: 'Asymptotically Sub-Minimax Solutions of the Compound Decision Problem' in J. Neyman (*ed.*): *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pp. 131–48.

LIKELIHOOD

The fundamental question about statistical inference is philosophical: what primitive concepts are to be used? Only two answers are popular today. Edwards is the first scientist to write a systematic monograph advocating a third answer.[1]

[1] Edwards, A. F. [1972]: *Likelihood. An Account of the Statistical Concept of Likelihood and its Application to Scientific Inference.* Cambridge: Cambridge University Press. £3.80. Pp. xiii+235.

'Orthodox' statistics admits only one primitive concept, namely physical probability, a propensity that betrays itself in stable long run relative frequencies. The statistician devises procedures for testing hypotheses and estimating parameters; he chooses among possible procedures by pointing out desirable properties that show up under repeated sampling. There is much infighting about what properties *are* desirable—unbiasedness, minimum variance, size, power, significance level and all that—but all these properties have to do with long run operating characteristics of statistical procedures. The orthodox statistician, be he a follower of Neyman or Fisher or whoever, will not usually allow you to measure the credibility of any particular estimate or hypothesis. He allows you to say only that this particular estimate was made by a procedure that has virtues that would show up under repeated sampling, or, for example, that this hypothesis is rejected by a procedure that mistakenly rejects hypotheses only one per cent of the time.

Bayesian statistics tells quite the opposite story. The only primitive concept is degree of belief, which arguably ought to satisfy the probability calculus. If you do have probabilistic degrees of belief in the propositions in some field of interest, there is a simple model of learning from experience that Bayesians find satisfying. Orthodox statistics is characterised by a host of locally applicable, often *ad hoc*, but notably ingenious procedures. A single, simple global analysis is the delight of the Bayesian.

Thomas Bayes's original paper was published in 1763. The use of repeated sampling properties in inference goes back to Jacques Bernoulli, published 1713. Between these dates J. H. Lambert may have toyed with a third basic concept, although he subsequently developed the theory of errors in what I would now call the 'orthodox' way. In 1777 Daniel Bernoulli brought this third alternative more into the open. To take his engaging example, we know that one of two archers has been firing at a target. One is an untalented novice, the other a master. We observe that the shots cluster around the bull. Who aimed? Knowing the abilities of the two men, we reason: if the novice fired, then something very unlikely must have happened, but if it was the master, an event of far greater probability has occurred. On this evidence, we strongly favour the hypothesis, that the master shot the arrows at this target. If we restrict 'probability' to the orthodox, physical, usage, we cannot say anything about the *probability* that this is the target of the master, not the novice. (We could if we knew that the two men tossed coins to decide who would shoot, but that is not part of our data.) We can, however, follow D. Bernoulli and compare what R. A. Fisher called the *likelihoods* of the two hypotheses. Likelihood is a sort of inverse of physical probability. The likelihood of the hypotheses h, in the light of data e, is the probability of observing e if h were in fact true. D. Bernoulli apparently advises us to prefer hypotheses of greater likelihood given the data.

Euler at once retorted that this advice is metaphysical, not mathematical. Quite so! The choice of primitive concepts for inference *is* a matter of 'metaphysics'. The orthodox statistician has made one metaphysical choice and the Bayesian another. D. Bernoulli appears to have been proposing a third. Likelihood is of course formally defined in terms of probability, but it is being offered as a primitive concept of inference; primitive in the sense that it is supposed to justify inference, and that its use is supposed to need no further justification.

As a primitive concept likelihood did not fare very well, although one finds it suggested in surprising quarters. Edwards reminds us that John Venn, a good frequentist, seems to be prepared to use likelihood as a primitive tool in assessing statistical hypotheses, and that F. P. Ramsey, who put the subjective theory on a solid modern footing, also, in a late paper, comes down in favour of likelihood. But for all these hints the idea of likelihood would have lain fallow had it not been for Fisher. And Fisher was completely ambivalent about the concept. In informative asides Edwards reminds us of this ambivalence and in an historical paper (submitted to *Biometrika*) he has expanded this theme into a thorough history of likelihood.

It is important not to confuse Fisher's 'method of maximum likelihood' with likelihood as a primitive concept. The former is a valued technique in estimation theory, and can be justified by a large number of long run sampling properties. Fisher himself discovered many of these properties, and made the method of maximum likelihood an integral part of orthodox statistics. Equally, of course, likelihood is a crucial quantity for the Bayesian, who learns from experience by multiplying prior probabilities and likelihoods. But for neither orthodox nor Bayesian statistician is likelihood primitive; it is only a quantity that turns out to be central in many calculations. Fisher, from time to time, wanted likelihood to do more.

Fisher had no use for Bayesian analysis unless one is confronted by that rare situation in which hypotheses of interest are themselves generated randomly from a chance set-up. So he could not, in general, speak of the probability of an hypothesis. But he urged, from time to time, that relative likelihoods of hypotheses form the natural way to indicate the relation between hypotheses and evidence. This idea is present in some of the great work of the early 1920s, and recurs somewhat nervously in the middle thirties; it comes out again in *Statistical Methods and Scientific Inference*, his final major contribution. Fisher even goes so far as to say that likelihood is much like Keynes's idea of logical probability, except that it is not subject to the unjustifiable addition law for probabilities: we cannot add likelihoods to get the likelihood of some disjunction of hypotheses. Harold Jeffreys once told Fisher there was nothing wrong with postulating likelihood as a basic axiom for inference, and as Edwards puts it, Fisher 'wistfully' contemplated that, but was never altogether sure it was the right thing to do. Edwards, in contrast, has no doubts.

Edwards's book is more of a sermon than an attempt to provide logical foundations for a new mode of argument. Himself a geneticist, he is appealing to fellow scientists to reason in a certain way. He gives plenty of attractive examples, for the theory is to be tested by having good consequences for science. Edwards clearly has a strong 'intuition' that likelihood is the right tool, but unlike philosophers who boringly go on about their intuitions, Edwards aims at establishing a coherent body of method that enables the scientist to analyse his data in a sensible way.

Concerning the philosophical question, 'What primitive concepts to use?', Edwards is at first orthodox. There is only one kind of probability, the kind that shows up as stable relative frequency. He grants, fleetingly, that Bayesians have arguments showing that if one has degrees of belief in every sort of proposition, or if one is made to bet on anything under the sun, then in coherence one's betting rates will be probabilities. But he retorts that he simply does not have degrees of belief in the genetical hypotheses he contemplates, and no one forces him to lay bets on which hypotheses are right. Indeed the statistical models used in genetics are not properly called true or false; they are more or less adequate and there is no sense in betting on their truth.

Edwards tells his colleagues not to bet and so not to be Bayesian. He asks them to assess the relation between experimental evidence and statistical models of interest. He will not accept procedures justified merely by their long run operating characteristics: he wants to know how *this* particular piece of evidence bears on *these* hypotheses offered by *this* working model. Scientific inference, he believes, has essentially to do with particular cases. So orthodox statistical procedures are to be rejected. Edwards is left with likelihood.

He finds it convenient to measure relative likelihoods of hypotheses by taking natural logarithms, and he calls (relative) log-likelihood the measure of (relative) support for hypotheses by data. So to compare the support that e furnishes for h_1 against h_2, compare the logarithm of the probability of getting e, according to h_1, with the logarithm of the probability of getting e, according to h_2. Logarithms are used because this makes support 'additive' in a certain sense. If we have two independent pieces of data bearing on some hypotheses, the log-likelihood of the two items taken together is the sum of the individual log-likelihoods. So although we cannot combine the support for different hypotheses, we can combine different pieces of support for the same hypothesis. This, says Edwards, is exactly what we want in science, for a disjunction of hypotheses is no hypothesis at all. I may contemplate the hypothesis that the refractive index of a crystal is r, and the hypothesis that it is r', but there is no scientific hypotheses to the effect that the refractive index is r-or-r'. Philosophers, however, know to their cost that there is no good way to distinguish 'real' hypotheses from 'manufactured' ones. One can think of real cases as well as logicians' tricks. Surely I can contemplate the hypothesis that a certain quantity has a distribution from a specified part of a family of distributions (normal or log-normal or whatever) without having much idea about, or even interest in, specific parameter points. Can I not then ask how well supported this 'composite' hypothesis is?

In the case of simple hypotheses there is still a question about the actual log-likelihood numbers. Do they mean anything? The orthodox statistician will say he does not know what to *do* with them. Probabilities and operating characteristics let you do all sorts of things. Given some utilities, they allow you to compute expected loss. But what can we do with likelihoods? Edwards's reply is two-fold. First, he does not want to do any of the things orthodox statisticians can do. He attaches no sense to a loss function over hypotheses in genetics. He wants to report evidence and show, in brief form, how it bears on the hypothesis under examination. All right, but what do Edwards's numbers mean? He has an unexpected but sensible answer. Use likelihoods and you will find out from experience what the numbers mean. It takes a while to learn, for example, what temperatures mean, and it is notoriously hard for Fahrenheiters to take in weather forecasts in Centigrade, but spend a summer in the South of France and you get to know what 30°C feels like. Numbers that basically record a ranking take a lot of getting used to. Indeed we could make Edwards's remark even about probabilities. Shortly before his death L. J. Savage was saying that we have only just begun to get a grasp of personal probabilities, and it might take several generations more before they were properly entrenched in our understanding. Whether or not he is right, it is clear that numerical probabilities have been radically extending their domain and their intelligibility. A meteorological classic from the early decades of this century tells the forecaster never to use the word 'probable' for it would be redundant in a weather forecast, all forecasts being merely probable. Nowadays the U.S. Public Weather Office seems incapable of uttering a sentence not qualified by a probability number. Savage, incidentally, thought the Weather Office simply did not know what its own probability percentages meant. Edwards could carry his fight into the enemy camp: if you use likelihoods in reporting experimental work, you will get to understand them just as well as you now think you understand probabilities.

Many philosophers will resent this kind of reasoning but I do not find it intrinsically disturbing. What does worry me is Edwards's faith that a given log-likelihood ratio will mean the same in any circumstance whatsoever. Grant, for a moment, that if the likelihood of h_1 on e exceeds that of h_2, then, lacking other information, e supports h_1 better than h_2. Now suppose the actual log-likelihood ratio between the two hypotheses is r, and suppose this is also the ratio between two other hypotheses, in a quite different model, with some evidence altogether unrelated to e. I know of no compelling argument that the ratio r 'means the same' in these two contexts. Physical probability is much better off. There is always the frequency interpretation telling us that if the probabilities of two unrelated events, in different chance set-ups, are the same, then the two events tend to occur equally often. That, I think, is the chief virtue of the frequency interpretation: it shows that different probabilities in different set-ups are commensurable. No non-Bayesian argument shows that likelihood ratios in different situations are always commensurable, that is, measure the same levels of evidential significance.

Indeed in artificial cases there seem to be positive counterexamples to unrestrained use of likelihood. A classic case is the normal distribution and a single observation. Reluctantly we will grant Edwards that the observation x is the best supported estimate of the unknown mean. But the hypothesis about the variance, with highest likelihood, is the assumption that there is *no* variance, which strikes us as monstrous. Edwards is a practical reasoner, and is inclined to disregard this case. If we do wish to fit it into the likelihood scheme of things, we must concede that as prior information we take for granted that the variance is at least w. But even this will not do, for the best supported view on the variance is then that it is exactly w.

For a less artificial example, take the 'tram-car' or 'tank' problem. We capture enemy tanks at random and note the serial numbers on their engines. We know the serial numbers start at 0001. We capture a tank number 2176. How many tanks did the enemy make? On the likelihood analysis, the best supported guess is: 2176. Now one can defend this remarkable result by saying that it does not follow that we should estimate the actual number as 2176, only that comparing individual numbers, 2176 is better supported than any larger figure. My worry is deeper. Let us compare the relative likelihood of the two hypotheses, 2176 and 3000. Now pass to a situation where we are measuring, say, widths of a grating, in which error has a normal distribution with known variance; we can devise data and a pair of hypotheses about the mean which will have the same log-likelihood ratio. I have no inclination to say that the relative support in the tank case is 'exactly the same as' that in the normal distribution case, even though the likelihood ratios are the same. Hence even on those increasingly rare days when I will rank hypotheses in order of their likelihoods, I cannot take the actual log-likelihood number as an objective measure of anything.

Edwards's book has many technical and practical suggestions beyond the scope of this journal. There is one 'philosophical' novelty to be remarked: the device of imaginary experiments. This has had some small play in Bayesian literature, and perhaps goes back to Laplace, but it is newly introduced into likelihood. Suppose I am thinking about some hypotheses and have not, as yet, got any new data. Still, I am not indifferent between the hypotheses; I have background information and prejudices. I realise, perhaps, that I regard h_1, in comparison to h_2, *as if* I had evidence e conferring a log-likelihood ratio r between these two hypotheses. This, then, is my 'prior support'. When I do get some real data, I can add the prior support and the actual experimental support to get a posterior support.

Edwards's very definition of prior support reads oddly. 'The *prior support* for one hypothesis against another is S if, prior to any experiment, I support the one against the other as if I had conducted an experiment . . .' (p. 36). In most of the book it is *data* that do the supporting, but here it is *me!* After I have done a real experiment, I am supposed to add 'my' prior support to the support furnished by the evidence; that looks dangerously like using arithmetic to add pints of milk to pounds of apples. Edwards notes that typically prior support counts little towards posterior support, so this may sound like the Bayesian situation with prior and posterior probability. But the Bayesian prior is the same kind of thing as the posterior, and moreover the prior is needed to get a posterior. Edwards's prior support seems to me a different thing from the support that evidence furnishes for hypotheses, and the latter kind of support does not require the former.

Edwards's book is nicely written, has a straightforward development, plenty of examples and instructive historical asides. It is not a book written for philosophers, but it is a book that those who care about probability ought to read. It gives a puzzling and possibly fundamental inferential concept a longer run than anything published before now. Statistical reasoning is not well enough understood by anyone, yet, and we need more fundamental concepts in the arena than is currently the case. I do not know how Edwards's favoured concept will fare. The only great thinker who tried it out was Fisher, and he was ambivalent. Allan Birnbaum and myself are very favourably reported in this book for things we have said about likelihood, but Birnbaum has given it up and I have become pretty dubious. George Barnard is the only worker who has consistently and persistently advocated and advanced a likelihood philosophy. I hope Edwards's book will encourage others to enter the labyrinth and see where it goes.

IAN HACKING
Cambridge University