and provided any procedure is interpreted under an adequate (conditional) probabilistic setup.

Thus I agree with Professor Lindley that the calculus of probability is a privileged instrument to operate with and to represent uncertainty, but I am not convinced that only probabilities conditional on available information are finally relevant. Sampling probabilities, such as significance levels, may also convey useful information precisely for being conditional on unobservable parameters; but they should clearly be properly interpreted and, in particular, not be confused with posterior probabilities; this later issue has been aptly argued in Berger and Delampady (1987)

In conclusion, I very much enjoyed reading Professor Lindley's fascinating exposition. I nevertheless stick to the idea that developing a proper understanding of both the sampling theory and the Bayesian paradigm is more appropriate than overdeveloping one and ignoring the other. I believe that this attitude is likely to be more fruitful for statistical practice and for understanding within the statistical community.

ADDITIONAL REFERENCES

DRÈZE, J. (1987). *Essays on Economic Decisions under Uncertainty.* Cambridge Univ. Press, New York.

DRÈZE, J. and MOUCHART, M. (1989). Tales of testing Bayesians. CORE Discussion Paper No. 8912, Université Catholique de Louvain, Louvain-la-Neuve, Belgique.

FLORENS, J. P. and MOUCHART, M. (1988). Bayesian specification tests. CORE Discussion Paper No. 8831, Université Catholique de Louvain, Louvain-la-Neuve, Belgique. To appear in *Contributions to Operations Research and Economics: The Twentieth Anniversary of CORE* (B. Cornet and H. Tulkens, eds.). MIT Press, Cambridge, Mass.

Rejoinder

Dennis V. Lindley

I am most grateful to the editors for inviting so many fine statisticians to comment on these lectures and to the discussants for raising so many important and interesting points. Where a point is mentioned by two or more, it appears under the first. This has the apparent difficulty that later ones appear to deserve a shorter reply.

COX

2. The Fisherian tradition (also mentioned by Barnard) is usually better suited to the treatment of scientific and technological data than that of Wald. The latter was discussed because the Wald lectures, on which this article is based, were given in the States, where Wald's ideas are more commonly encountered than in Britain. A critique of the Fisherian view has been given by Basu (1988).

3. A common objection to the Bayesian view is "where did you get that prior?" (Section 1.3); hence the emphasis on elicitation. A careful reading of Jeffreys will show that he often does not use flat priors (for example, in hypothesis testing). More recent work has shown multivariate, flat priors to be unsatisfactory. Why should personal judgment be left qualitative? The failure to quantify can lead to imprecision and vagueness.

4. Cox agrees with de Finetti in not liking axioms. The key question is surely our attitude to uncertainty: how are we to appreciate an incompletely understood world? A basic assumption is not that all probabilities are comparable, but that all uncertainties are. This assumption seems reasonable until someone can produce criteria that divide uncertainties into two or more types. So far as I am aware, this has not been done.

6. Temporal coherence has received relatively little attention. Some additional assumption seems called for. One approach is to recognize that probability statements typically contain *three* arguments and can be written $p(A \mid B:C)$, read as the probability of A conditional on knowing B and supposing C. The distinction between B and C is that the former contains known events, the latter events supposed to be true. Thus, in the distribution over sample space, a parameter would usually belong to C since its value is unknown. Temporal coherence requires an extra axiom, $p(A \mid B:CD) = p(A \mid BD:C)$. This says that as D passes from merely supposed, to experienced or known, the probability does not alter. With the axiom one can write $p(A \mid B:C)$ as $p(A \mid BC)$ and the distinction between supposition and fact is irrelevant. (Mouchart, in his contribution, reminds us that Pratt, Raiffa and Schlaifer, 1964, similarly felt the need for an extra axiom.) If the probability does change in practice, this is because more was experienced than merely D and that this additional experience was not part of the conditioning in the original probability statement.

The example is similarly handled using conditioning. The original probability over 10 and -20 was conditional on θ taking one of these values. If $\theta = 0.1$, or other values, are contemplated, then the conditioning event changes. There is no need to introduce epsilon. Probability is a function of two arguments and the neglect of the second can lead to misunderstandings and sometimes error.

7. I am unable to understand why the Bayesian argument ignores (a). Suppose that there are various sources of information concerning the efficacy of a drug, some from experiments in the States, some from Europe. There may be discrepancies due to geography. If so, a model introducing nuisance parameters to allow for the geography would be contemplated. The difficulty then is to remove the nuisance parameters. The Bayesian approach uses integration.

It is a pleasure to end on a point of agreement: "the key issue . . . is the judicious choice of questions asked and . . . the choice of model." The Bayesian view is that the key question is what is your uncertainty about the quantity of interest, and that the model should be entirely probabilistic. There is no need to live in a complicated world of eclecticism and adhoceries: the unifying principle is probability addressed to the key question of uncertainty.

BARNARD

The argument in the sixth paragraph that GFNP is concerned to design experiments, leaving the client to supply background information, is surely wrong because background information must be used in design. What else is there? Although the article discussed only post-data analyses, the Bayesian view encompasses experimental design using the same probabilistic expressions of uncertainty in the design and in the subsequent inference.

It is a pleasure to join in the praise of the likelihood function. It was a triumph of Fisher to recognize its importance and brilliant of Barnard to appreciate the likelihood principle. From an entirely different tack, the Bayesian view comes to agree with them. To my mind, it is one of the tragedies of recent years that our profession has failed to accept the principle, especially after the publication of Berger and Wolpert (1984). Editors ought to insist that any submitted paper containing methods that violate the principle should explain why it is unsound.

Where the likelihood principle alone is inadequate is in dealing with nuisance parameters. The discussions that Barnard gives apply to special cases, like the existence of pivots, or are approximate, as with contingency tables. These can be viewed as almost Bayesian and sit comfortably within the canon. But why use these special devices when a general method is available? Of course, as Barnard shows, the recognition of a pivot can simplify the argument and also allow for an easier treatment of robustness issues. Such ideas have limitations though, as is seen when issues like "nothing else is known about [a parameter]" arise. Despite the substantial effort that has been put into finding a precise description of ignorance, none has been found outside the finite case. The many counter-examples testify to the elusive nature of ignorance. My own view is that if you understand the role of the parameter then, by definition, you are not ignorant.

The discussion around the Michelson–Morley experiment is enormously clarified for me by the distinction between models and theories (Section 2.5). It is easy to think of an alternative model in the experiment. It is far from easy to construct a theory that will incorporate that model as a special case. Thus the comparison by means of likelihoods derived from the model is still possible, even in the absence of a theory. Repeatability is important because it enables uncertainty to be reduced, expressed by a change in the probability distribution. However, notice that it is rare, and even undesirable, to have pure repetition. There is usually some fundamental change in the later studies. For example, better equipment may be employed, or if, as with cold fusion, another explanation is possible, the experiment may be redesigned to investigate it. Statisticians love repetition (exchangeability), but I suspect this is because they are wedded to frequency ideas and find difficulty, as Cox does, in experiments that are not mutually consistent.

BERGER

1. I wish it was possible to agree with the statement that "[Bayesian statistics] is much easier to understand and yields sensible answers with less effort." Mostly this is true, but often not. I have spent some time studying inference concerning Hardy–Weinberg equilibrium (Lindley, 1988). The sampling-theory solution due to Haldane is extremely simple to use. The only difficulty is showing how good it is. This is an issue that need not bother practitioners. The only Bayesian approach I know involves careful consideration both of exactly what is being tested and the alternatives. The geneticist would benefit from such considerations, but he will need convincing that the complexity is worthwhile before abandoning Haldane's brilliant and simple solution.

2. The article is defective in not saying precisely what is meant by coherence. Loosely it is that our separate judgments should fit together, or cohere, just as Berger's E and E' should. Precisely, it has to be formulated in terms of assumptions (like Savage) or through a basic principle (like avoidance of Dutch books, following de Finetti). Either way the complete

probability structure is obtained. In particular, the conditioning principle is included, if only as a consequence of the likelihood principle. Statisticians who use Bayes rule and prior information but base their evaluations of accuracy on frequency ideas would be incoherent because in two situations with identical likelihoods they would usually have different evaluations, this in violation of the principle.

3. In the present state of development of Bayesian statistics, I view methods that deal with classes of probabilities and those that consider errors in assessment (as in the article) to be alternative methods of tackling the same problem. Time will tell which is the more effective. A difficulty with the class approach is that it often uses extreme values. Thus it is common to find the worst, in some sense, amongst a class. Then it can happen that the worst is an extreme distribution that one would hardly contemplate. One thing Bayesian theory teaches us is the value of averaging in the use of expectations. Maximum likelihood should be replaced by average likelihood. It therefore seems sensible to average over a class of probabilities, so obtaining another probability. This is what happens as far as the second order in the methods of Section 6. The argument that the paradigm "surrenders the axiomatic high ground" would equally apply to Euclid or Newton. Newton's theory was accepted long before the third significant figure in the distance between America and Europe could be determined. And contrary to what Berger says, there is a theory that helps one "concentrate measurement efforts" to get good triangulations. Certainly some probabilities are long going to resist accurate determination but others have already yielded useful results, for example, in actuarial science.

LEHMANN

1. The adjective "broadly" is there precisely because of the existence of some Bayesians in the department at Berkeley.

2. The Bayesian paradigm is restrictive in that it demands that all uncertainty be expressed probabilistically. But it is undemanding in that this is the only restriction; in particular, the probabilities can be evaluated in any manner. Thus, in the evaluation of the likelihood function it may be expeditious to contemplate a sample space and a class of probabilities over it. In no way does this offend Bayesian principles. The models of this type commonly used in statistics are correct and valuable. The point being made is that once the data are to hand, only the likelihood (and prior) matter and the sample space by which one ascended can be discarded.

A role of parameters is to simplify the probability structure. (How right Lehmann is to remind us of Fisher's introduction of these.) The reality is data, past and future, and the (Bayesian) statistician's task is to attach probabilities to them. This is difficult, but the introduction of parameters that yield iid observations makes an enormous simplification. Thus the standard models are effective and involve $p(x \mid \cdot)$. Models are important for Bayes and Berkeley alike however much data analysts might eschew them.

3. It is a pity that Lehmann and van Dantzig feel the need to invoke emotive phrases like "Holy Land of Statistics" or "Statistical Priesthood." Perhaps it reflects on their difficulty in replying to a logical argument by logical reasoning. Bayesian statistics is no more Utopian than Newtonian mechanics and will one day be found to be inadequate by some future Einstein. In the meantime reasoning, not religion, suggests that it is the best available.

BERNARDO

1. The practical example of the value of extending the conversation is excellent. It demonstrates the merit of using well-determined values, $p(j)$, and also of introducing another variable, j.

2. Perhaps because of the failure of Jeffreys's method of determining non-informative priors, and my own failure to amend it, I have been rather opposed to their use. But the more recent work of Bernardo and the use made of them by he and Berger, make me change my mind. The mistake is to think of them as representing ignorance. Rather they provide a reference distribution, corresponding to little knowledge, that is useful in communication between personal probabilists. De Finetti has argued that each problem should be considered on its merits. This is fine, but some problems almost repeat and standard forms of likelihood and prior are useful. Berger puts it nicely when he says a subjective prior is better than a non-informative one, but the latter is "very good."

Bernardo rightly reminds us that statistical problems are ultimately decision problems. The statistical literature is full of complaints about bad statistical practice. But how can a tail-area significance test be used to decide? Of what practical value is a confidence interval? How can you act on the basis of a likelihood function? Only the Bayesian approach provides a reasonable paradigm for the single decision-maker.

MOUCHART

The editors have only given me three days in which to reply and in this time there has not been the opportunity to see Drèze's work. It would surprise me if Savage's model were not appropriate but a theory develops or fails by being tested and I will study the reference.

The difficulty of incorporating the costs of elaboration and of operation of the model is a very real one. Numerical integration can be expressed as a decision problem, the values of the function being uncertain. It is not difficult to write down a Bayesian solution for the optimum points at which to evaluate it. However, this solution involves a numerical integration substantially harder than the initial one.

I cannot agree that significance levels are ever necessary. They sometimes appear to be valuable because we all have occasional trouble distinguishing $p(A \mid B)$ from $p(B \mid A)$. Here A refers to the parameter and B to the sample space. We want the former but, in the test, make do with the latter. But they ultimately fail because of the violation of the likelihood principle. The objection that likelihoods do not always exist, for example, because of the lack of alternative hypotheses, is met by using models (rather than theories) and these are usually easy to produce as mentioned in the reply to Barnard.

FRENCH

Utility and probability surely measure different types of judgment but French's discussion does alert me to a problem; namely how can they separately be elucidated? Is there any way in which a probability can be assessed without involving some utility considerations? For example, if a scoring function is used, how can one know that the assessor's utility is linear in the score? Can a utility be separated from a probability? The discussion of Allais's paradox in Section 4.4 suggests not. The mathematics always involves the product of the utility and the probability, thereby suggesting difficulties of separation. In one of my first papers (Lindley, 1953), the development was entirely in terms of this product, called a weight, never separating the weights into their component parts.

French objects to the analogy with Euclidean geometry partly on the grounds that "the earth is not a perfect sphere." But geometry does not demand this and the theory is happy with reality. He also draws our attention to hills and valleys. Decision analysts need to consider them if the variation is important. There are two points about Cleethorpes. First, the probability should be conditional on knowledge that the subject possesses, knowledge which might change as further thought is devoted to the problem. Second, true values seem to be useful concepts within the mathematics, even if their reality is debatable. Parameters and true values are similar in that their introduction substantially simplifies the mathematics (see the discussion of Lehmann). Often it is convenient to evaluate $p(\theta \mid x)$ because it will then be available for any future prevision about data that depends on θ.

I cannot agree that "all statisticians agree on the use of probability to model uncertainty." Berkeley does not measure the uncertainty of a hypothesis by the probability of the hypothesis. Nevertheless one can completely agree with French that other measures of uncertainty should be avoided and that probability is the only serious candidate.

KADANE

One of the many mistakes I have made in my life was to underestimate the computational difficulties in the Bayesian approach. I failed to appreciate that even a reasonably complete theory can be difficult to implement and that substantial effort would have to be directed to the numbers. Kadane is so right to emphasize this and provide an adjunct to the lectures that makes them more balanced. Where do the numbers come from? How do we handle them? How sensitive are the results to fluctuations in their values? What is the best way to obtain good numbers? All these are important questions and their resolution is vital for the success of the Bayesian approach.

EPILOGUE

The discussion has been interesting not only for the excellent points made but also for what has not been said. As Holmes observed, the fact that the dog did not bark is important. This feature has been present in other discussions of Bayesian ideas or of the likelihood principle and a few moments exploring the omissions may not come amiss.

The modern foundational approach to Bayesian statistics consists in tight, logical arguments. These take many and varied forms but all lead essentially to the laws of probability with minor, though sometimes important, variations. It is surely reasonable to expect that any critical response would take parts of the logic and criticize them, just as geometers considered the parallel postulate of Euclid. With a few honourable exceptions, such as Allais's, this does not happen. Objections take the form of generalities. We are told the paradigm does not address the right questions. If so, let us see these questions put into a logical framework. What will that look like? Faced with that, even logic is derided and various terms like flexibility invoked to support the collection of varied, popular statistical practices.

The sampling-theoretic procedures have been investigated by writers and almost all of them have been found to contain flaws, mostly easily conveyed through counter-examples. These counter-examples are rarely mentioned by non-Bayesians. I can still remember the gasp of astonishment that a Berkeley audience expressed when presented with one (by Lehmann, one of the few who have seriously addressed them). Barnard mentions a modified definition of ancillarity that avoids the startling counter-example of Basu's. But such serious attempts to face up to the defects of accepted practice are rare. People are generally ignoring the construction of the Bayesian edifice and the knocks given to their own. It is this silence that Holmes might have used as evidence in favour of the coherent position.

ADDITIONAL REFERENCES

LINDLEY, D. V. (1953). Statistical inference (with discussion). *J. Roy. Statist. Soc. Ser. B* **15** 30–76.

LINDLEY, D. V. (1988) Statistical inference concerning Hardy-Weinberg equilibrium (with discussion). In *Bayesian Statistics 3* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.) 307–326. Oxford Univ. Press, Oxford.