has approximately $D(a)/a^2 \sim 324$ and where C is a universal constant (≤ 10 for large $D(a)$).

One would then expect that, for a loss function $H^2(t, \theta)$, and for prior measures μ that are sufficiently well spread out, the Bayes estimates β_n would satisfy a similar inequality: $E_\theta H^2(\beta_n, \theta) \leq C'D(a)$. This is indeed the case. However, we could not find measures μ that are sufficiently well spread out except under a severe growth restriction on $D(\tau)$ as $\tau \to 0$. Roughly, the growth restriction is that $D(\tau)$ increases slower than $\tau^{-1/3}$ as $\tau \to 0$. This rules out interesting cases, such as the case where Θ is the set of bounded densities satisfying a Lipschitz condition on the unit square of the plane. The nonparametric sets used by Diaconis and Freedman have dimensions that increase very rapidly as $\tau \to 0$, even if the distances used are much weaker than our H. Most small open sets have positive but essentially negligible probabilities.

To obtain better results, it seems necessary to take into account features of the statistical problem that are not summarized by the distance H. Which features are most important is presently a matter of conjecture. Here, Diaconis and Freedman suggest a direction of study that may be very important: They investigate the derivative of the posterior measure viewed as a function of the prior measure. Now, let $\mu \cdot P$ be the marginal measure $\int P_\theta \mu(d\theta)$, let $\mu \otimes P$ be the joint distribution, and let K_x be the conditional distribution of θ given x. Then, with the present symbolism

$$(\mu \cdot P) \otimes K(\mu, P) = \mu \otimes P.$$

This relation can be differentiated not only in μ but also in P. For instance, retaining only first order terms in ε, one would have

$$(\mu \cdot P) \otimes \{K(\mu, P + \varepsilon\Delta) - K(\mu, P)\} \sim \varepsilon\{\mu \otimes \Delta - \Delta \otimes K(\mu, P)\},$$

a relation analogous to the one given by Diaconis and Freedman. It may be feasible from such relations to find out which features of μ or $\{P_\theta: \theta \in \Theta\}$ influence the posterior distributions and the attached risks. However, as far as we know the subject has not yet been studied in sufficient detail.

Perhaps my formerly Bayesian colleagues will tell us in the near future what pairs (μ, P) are "safe" and what pairs are bound to give trouble.

Department of Statistics
University of California
Berkeley, California 94720

Dennis V. Lindley

Somerset, England and Monash University

My own view of statistics is that it is a way of studying some aspects of the real world, namely the uncertainty present in any study, and of expressing my beliefs about the world. The subject is not primarily mathematical but mathematics plays an essential role because it enables me to pursue the logical consequences of beliefs and see whether they conform with other beliefs. It is perfectly possible for the logic alone to change my beliefs. For example, at the moment of writing my belief that the 24th digit in the decimal expansion of π is 4 is expressed by a probability of 0.1. Were I to do the mathematics, or accept the mathematics that others have done and consult a book, I would find that the digit is 3 and my probability is now 0 (or at least very small: The book might be in error or I could have erred in reading it). The most impressive example of logic changing beliefs is the work of Ramsey, Jeffreys, Savage, and de Finetti in demonstrating that beliefs need to be measured probabilistically and not, for example, by significance levels.

Now in the present paper (referred to as DF) we have some impeccable logic that shows that in certain circumstances the Bayes estimate will be inconsistent. Just as the book changed my opinion of π, so DF changes my beliefs about the estimation of a location parameter. Before discussing this let us clarify two points. In applying mathematics to the real-world problems of statistics it is always necessary to be reasonably sure that the mathematical modelling has been sensibly done. (Some might object to Ramsey's work on these grounds.) The modelling in DF does seem reasonable to me and the results cannot be dismissed on these grounds. Secondly, I do object to the use of Bayes estimates. These are just a carryover from the inept modelling of sampling-theory statistics. The Bayes "estimate" of θ is the probability distribution of θ given the data. As far as I can see this does not affect the conclusions of DF since their theorems relate to μ_n and not $\hat{\theta}$ (Equation (1.1)).

So what am I to make of the mathematical results of DF? Clearly they change my beliefs in some way, but how? One thing I could do is to change my prior beliefs and not use a Dirichlet with Cauchy measure. Jeffreys (1967) does something like this in a different context. In Section 5.2 he notes that a normal prior would lead to posterior views that are unacceptable to him: so he uses a Cauchy form and all is well. In DF I could replace the Cauchy by a normal. But it may be that the Cauchy form does adequately reflect my opinions so that the inconsistency persists. Now the result of DF tells me to beware of $\mu_n(\theta)$, at least in certain cases. But presumably if h really had the eccentric trimodal form of Figure 1, the empirical distribution function would reveal this as a serious possibility. Looking at that function I would loosely argue something like this. DF warned me about these trimodal fellows and yes, $\mu_n(\theta)$ does keep oscillating between the left- and right-hand values, so I had better change my view and think that the location is at the central mode.

It is not clear to me how logical results should change my beliefs. Bayes showed us how to change with data but is there some sensible way to react to mathematics? (With π it was easy.) Here is a very simple example of the problem. I am considering two events A and B and after reflection assign probabilities $p(A)$ and $p(B)$, perhaps 0.6 and 0.5. Now DF comes along and demonstrates that A and B are exclusive, a fact I had not known. What are reasonable values of $p(A)$ and $p(B)$ now? One way suggested in Lindley et al. (1972) is to think of $p(A)$ and $p(B)$ as, in some sense, assessments, subject to error of "true" values $\pi(A)$ and $\pi(B)$. The observation of DF amounts to saying $\pi(A) + \pi(B) \leq 1$ and the space of their values can be restricted accordingly. But even I am not quite happy with this. The problem arises in sampling-theory statistics. Lindsay (1980) suggests estimating a binomial parameter, p say, by r/n, in the standard notation. But a little calculation in the problem shows that $p = \frac{1}{2} - 2\theta^2$: What is the estimate now (especially when $r/n > \frac{1}{2}$)?

There is a further aspect to the results of DF: They point out that θ can be estimated by using the median. They say "Bayes estimates do worse than available frequentist procedures." Is this justified? There are two possibilities: either the median is a Bayes estimate or it is not. (By a Bayes estimate here I mean for a *fixed* prior for *all* sample sizes. Some frequentist procedures, like significance tests, are Bayes for each n but the prior to make them Bayes has to change with n. The Bayesians of DF are not allowed this luxury.) If it is a Bayes estimate the quoted claim is false. If not, then what is the frequentist doing using an inadmissible procedure? Could not the coherent Bayesian make the median user lose money for sure?

As a paper about statistics—and although there is understandably no hint of a real-world usage, I take it we are discussing it as a statistical paper—it does not lessen my respect for the Bayesian argument but it does reinforce doubts about how a Bayesian should react to logical deductions, as distinct from data. The logic of DF therefore has important, and to me, unresolved consequences.

I conclude with a few miscellaneous remarks.

(a) The Dirichlet prior is unacceptable to me because it fails to incorporate the positive correlation that I feel between adjacent, nonoverlapping intervals. As a result the posterior is insufficiently smooth. Do the inconsistency results persist with some smoothing present?

(b) There is a strong reason for Bayesians being interested in frequentist results because the latter are useful in experimental design (preposterior analysis). Before the data are to hand they are random and accordingly governed by probability laws; the likelihood principle does not obtain and the sample space is relevant.

(c) It is easy to produce examples of Bayesians (and others) being misled. Let X_i be i.i.d. $N(\theta, 1)$ and let sampling continue until the hypothesis that $\theta = 0$ is rejected at the two-sided 5% level; that is, until $|\overline{X}| > 2\sigma/\sqrt{n}$, $\overline{X}$ being the mean of a sample of size n. This is certain to happen. Let θ have a uniform prior. Then $p(\theta < 0|\overline{X}, n)$ must always be less than $2\frac{1}{2}\%$ or greater than $97\frac{1}{2}\%$, since θ is $N(\overline{X}, \sigma^2/n)$.

(d) An earlier version of this paper was given by Freedman at the 1983 IMS meeting in Toronto. I did not attend the meeting but for months afterwards I had people coming up to me with undisguised glee telling me about the paper and implying that Bayesianism was now dead. May I remind any who think this that *all* frequentist procedures have counterexamples far simpler and far more devastating than any this paper contains. I do not remember these being discussed in the IMS journals: How about it, editor? I collected a few together in Lindley (1972).

(e) My first, quick reaction to this paper was to dismiss it as modern mathematics out of control again. This is grossly unfair. The authors model commonly occurring situations in apparently sensible ways and produce unexpected results. It has given me much to think about and will continue to do so after the deadline for submission of this comment has passed. Diaconis and Freedman have done us a service in exploring the consequences of apparently innocuous assumptions so carefully.

REFERENCES

JEFFREYS, H. (1967). *Theory of Probability.* Clarendon Press, Oxford.

LINDLEY, D. V. (1972). *Bayesian Statistics, A Review.* SIAM, Philadelphia.

LINDLEY, D. V., TVERSKY, A. and BROWN, R. V. (1972). On the reconciliation of probability assessments. *J. Roy. Statist. Soc. Ser. A* **142** 146–180.

LINDSAY, B. G. (1980). Nuisance parameters, mixture models, and the efficiency of partial likelihood estimators. *Philos. Trans. Roy. Soc. London Ser. A* **296** 639–662.

2 PERITON LANE
MINEHEAD
SOMERSET TA24 8AQ
ENGLAND

REJOINDER

P. DIACONIS AND D. FREEDMAN

Stanford University and University of California, Berkeley

Introduction. We would like to thank the discussants for their careful work. For context, we summarize our position.

(a) As a team, our motives are mixed to an unusual degree, because we differ on many issues in foundations, including the interpretation of some of our results. However, we are unanimous that the mathematics in our paper should be of interest to Bayesians, ex-Bayesians, and never-Bayesians alike.

(b) Frequentists can use the Bayesian approach, like maximum likelihood or optimality, as a powerful heuristic engine for generating statistical procedures. No such engine is foolproof, so you should always look to see how well the procedure is going to do. Even the crustiest subjectivist ought to follow this advice, when the prior is only an approximation (and possibly quite a crude one, chosen for computational convenience) to the true subjective belief. Besides its practical importance, checking operating characteristics is good, clean mathematical fun.

(c) Pitfalls in the classical approach are well known; those in the Bayesian approach perhaps less so. We have given some examples where plausible applications of Bayesian technique lead to disaster. It is particularly easy to lose your way in high dimensional parameter space.

(d) We view consistency as a useful diagnostic test. If your procedure gives the wrong answer with unlimited data, probably you will not like it so well with a finite sample either.

(e) We show how putting conditions on the underlying model and modifying the prior can sometimes rescue Bayes procedures. As a general heuristic device