GOOD, I. J. (1983). *Good Thinking: The Foundations of Probability and Its Applications*. Univ. Minnesota Press, Minneapolis.

HUBER, P. (1973). The use of Choquet capacities in statistics. *Bull. Inst. Internat. Statist.* **45** 181–191.

LEAMER, E. E. (1978). *Specification Searches*. Wiley, New York.

MORRIS, C. (1983). Parametric empirical Bayes inference: Theory and applications (with Discussion). *J. Amer. Statist. Assoc.* **78** 47–65.

POLASEK, W. (1984). Multivariate regression systems: Estimation and sensitivity analysis of two-dimensional data. In *Robustness of Bayesian Analyses* (J. Kadane, ed.). North-Holland, Amsterdam.

DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907

MURRAY K. CLAYTON

University of Wisconsin

The two papers by Diaconis and Freedman which are under discussion contain a series of interesting and nicely presented results. The philosophical issues which they raise are thought-provoking and merit attention. Their papers also give a useful review touching on a number of topics of interest to frequentists and Bayesians.

For simplicity, in the ensuing comments I shall refer to Diaconis and Freedman (1986a) as DFa and Diaconis and Freedman (1986b) as DFb. My comments touch on three topics: the technical aspects of DFa, the philosophical implications of the results in DFb, and the extension of the "what if" method in DFb to Bayesian robustness.

The model (1.1) of DFa and the accompanying priors seem innocuous, and it is somewhat disconcerting that they can lead to inconsistency. Theorem 1 of DFa says that the posterior for θ will fail to converge even though h has a global maximum at 0. Theorem 3 states that using a symmetrized prior might not help; we can even get the posterior law of the data wrong. On the other hand, perhaps the consoling message from DFa is that if $\log \alpha'$ is convex, then in the setting of Theorem 1 the posterior for θ *will* converge. Less helpful is the fact that the posterior will converge if the (unknowable) density h is strongly unimodal.

The discretization results of Section 4 of DFa can be used to approximate the solutions to decision problems in the undominated case. In Clayton (1985), I used a form of discretization with a Dirichlet process prior to approximate the worth of optimal rules for a sequential problem. I conjectured in that paper that discretization could be used to construct nearly optimal rules. (The construction of *optimal* rules is practically impossible unless the Dirichlet parameter has a finite support.) It seems possible to use the results of Section 4 of DFa to prove that conjecture.

How important is this issue of inconsistency to a Bayesian? I think Diaconis and Freedman are right in DFb to consider separately the classical and subjective Bayesians, even though many Bayesians have the characteristics of both groups.

To a classical Bayesian, a consistent Bayes estimate means that the Bayesian will eventually discover the "true" parameter value, and so the Bayesian and frequentist will eventually agree. This seems consoling to those ill-at-ease with the Bayes/non-Bayes controversy—we might use different methods, but we eventually uncover the same truths.

Why should a subjectivist be concerned with inconsistency? Theorem 3 of DFb says that if a Bayesian "A" is consistent, then from a Bayesian "B's" point of view A and B will eventually agree. There are situations, however, where this is unsuitable. For example, suppose A and B are witness to some coin tossing. Bayesian A is firmly committed to the belief that all coins are fair, and so uses a prior $\delta_{1/2}$ for θ, the probability of heads. B is firmly committed to the belief that coins are never fair, and uses a uniform prior on $[0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$. Both A and B will use Bayes theorem to coherently update their priors as they see data, but they will never agree, nor should they.

Such failures of opinions to merge do arise, at least informally, in discussions about issues such as the afterlife, a flat earth, the Bermuda triangle, the cause of mass extinctions, and so on.

In the spirit of the "what if" method, the subjectivist should be interested in the effect the prior has on inferences, and should therefore be interested in its possible inconsistency. However, while the classical Bayesian would probably never want to use an inconsistent prior, the subjective Bayesian might very well choose to use such a prior.

I agree with Diaconis and Freedman that the "what if" method is a useful technique in considering a prior; I suggest that it provides a method of assessing the robustness of the analysis to the choice of prior. As Berger has suggested (Berger, 1984), it is difficult to specify priors exactly and so we should be aware of how changes in the prior will affect our inferences. If a small change in the prior results in a small change in the posterior, then this indicates that an exact specification of the prior is not critical. On the other hand, if a small change in the prior results in a large change in the posterior, then the data have little to say relative to the information in the prior. Presumably in this situation one would want to be more careful about the choice of prior.

While Berger (1984) tends to look at gross changes in the prior, I would put the emphasis on examining "small" or "local" changes in the prior. The effects of small changes of this sort are measured by the derivative of the posterior with respect to the prior, or in the notation of DFb, $\dot{T}_\mu$. Often we will be less interested in the entire posterior than the corresponding Bayes rule, and in that case we will prefer to look at $\dot{M}_\mu$. We can go further: In some settings a small change in the prior might result in a big change in the Bayes rule M, but this might not be important if the accompanying change in the Bayes risk, R, is small. This leads us to look at the derivative of the Bayes risk with respect to the prior, $\dot{R}_\mu$. $\dot{T}_\mu$ and $\dot{R}_\mu$ give "local" measures of Berger's "posterior robustness" and "procedure robustness," respectively (Berger, 1984).

How well does this approach work? In the example in Section 4 of DFb the interpretation of $\|\dot{T}_\mu\|$ is pleasing: The posterior of μ is most influenced when x is far from μ_0, relative to $\sigma^2 + \sigma_0^2$. This suggests picking a prior for which σ_0^2 is large. It is tempting to try to find "robust" priors for which $\|\dot{T}_\mu\|$ is small. However, $\lim_{\sigma_0^2 \to \infty}\|\dot{T}_\mu\| = \infty$ for μ_0, x, and σ^2 fixed, which leads to picking a prior with small σ_0^2. If my calculations are correct, $\|\dot{M}_\mu\|$ behaves similarly in this regard.

A possible complaint about using $\dot{T}_\mu$, $\dot{M}_\mu$, or $\dot{R}_\mu$ as measures of robustness is that only the prior is being called into suspicion; the likelihood is assumed fixed. One approach to this problem is to use a large class, C, of distributions for the likelihood: Box and Tiao (1973) give an example where a class of exponential power distributions is used instead of a normal likelihood. There is another approach which leads to a very large class C.

If we follow de Finetti (1975), then our efforts should concentrate on modeling observations, not parameters. Specifically, we should focus on P, the joint measure for the observable data $X_1, X_2, X_3, \ldots$. Suppose data $X_1, X_2, \ldots, X_n$ are collected and a Bayes rule M_n is formed. I would argue that from the robust Bayesian viewpoint it is appropriate to look at the derivative of M_n with respect to P. (Depending on how we define the neighborhoods of P, this could lead us back to $\dot{T}_\mu$.) Generally, dM_n/dP could be very difficult to compute. A compromise, which is not likelihood dependent and corresponds to a very large C, uses the Dirichlet process to describe the distribution P. The specific P chosen is determined by the Dirichlet parameter α. In keeping with the "what if" approach, we can ask how the Bayes rule is affected by a small change in α. To be more specific, let us suppose that $X_1 = x_1, \ldots, X_n = x_n$ are observed and we want to predict X_{n+1} with squared-error loss. The resulting Bayes rule is $\mu = M/(M+n)\mu_F + \Sigma x_i/(M+n)$ where $M = \alpha(\mathbb{R})$, $F(\cdot) = \alpha(\cdot)/M$, and $\mu_F = \int x\,dF$. We may compute the Gâteaux derivative, $\dot{\mu}(\alpha, \beta)$, of μ, with respect to α in the direction of the measure β. To keep the example simple, suppose $\beta(\mathbb{R}) = M$, $G(\cdot) = \beta(\cdot)/M$, and $\mu_G = \int x\,dG$. Then (Serfling, 1980)

$$\begin{aligned}\dot{\mu}(\alpha, \beta) &= \frac{M}{M+n} \lim_{\varepsilon \downarrow \infty} \frac{1}{\varepsilon}\left[\int x\, d\left[(1-\varepsilon)F + \varepsilon G\right] - \mu_F\right] \\ &= \frac{M}{M+n}(\mu_G - \mu_F).\end{aligned}$$

This suggests that, "locally," the only critical aspect in the mispecification of α is the mispecification of the mean of an observation. $\dot{\mu}(\alpha, \beta)$ also has the appealing feature of being small when n is large. When $G = \delta_z$, $\dot{\mu}(\alpha, \beta)$ differs from the influence curve for the mean by a multiplicative constant.

This approach is not easily extended to a model where P results from a mixture of Dirichlet processes, or if another "nonparametric" prior such as the tail-free prior is chosen. I am therefore eager to see the results Diaconis and Freedman obtain for the Gâteaux derivative in the undominated case.

Incidentally, the interpretation of $\|\dot{T}_\mu\|$ that follows from Theorem 4(b) of DFb is particularly appealing: A measure of sensitivity to the prior is obtained by looking at the ratio of objectivist likelihood to $D(\mu)$. Berger and others (see Berger, 1984, page 95 for references) have discussed similar uses of $D(\mu)$ and the likelihood for assessing model adequacy.

Finally, a query: In DFb, Diaconis and Freedman use the past tense in describing themselves as subjectivist and classical Bayesians, respectively. How do they describe themselves now?

REFERENCES

BERGER, J. O. (1984). The robust Bayesian viewpoint. In *Robustness of Bayesian Analyses* (J. Kadane, ed.). North-Holland, Amsterdam.

BOX, G. E. P. and TIAO, G. C. (1973). *Bayesian Inference in Statistical Analysis.* Addison-Wesley, Reading, Mass.

CLAYTON, M. K. (1985). A Bayesian nonparametric sequential test for the mean of a population. *Ann. Statist.* **13** 1129–1139.

DE FINETTI, B. (1975). *Theory of Probability* **2**. Wiley, New York.

DIACONIS, P. and FREEDMAN, D. (1986a). On inconsistent Bayes estimates of location. *Ann. Statist.* **14** 68–87.

DIACONIS, P. and FREEDMAN, D. (1986b). On the consistency of Bayes estimates. *Ann. Statist.* **14** 1–26.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

DEPARTMENT OF STATISTICS
UNIVERSITY OF WISCONSIN
MADISON, WISCONSIN 53706

A. P. DAWID

University College, London

The mathematical beauty and tractability of the Dirichlet prior render it almost irresistibly seductive. But beware! Rocks and shipwreck await the poor Bayesian navigator captivated by its siren song. Brown (1976) shone a little light on these murky waters. Now Diaconis and Freedman deserve the gratitude of all explorers for illuminating some of the more treacherous obstacles to a smooth passage.

Beyond these specific warnings, what broader morals are to be drawn? In view of the fact that, generically, the pair (θ, μ) is inconsistent, it is not really surprising that the authors can find such a pair. What I find far more surprising is the existence of priors μ (e.g., tail-free) which *are* consistent at each θ. Perhaps this is only possible because of the rather weak definition of consistency employed. Nevertheless, it is an important property, and one which demands further characterization.

Choosing a prior for an infinite-dimensional parameter space is always going to be problematical, and any accessible prop (such as consideration of imaginary results) should be grabbed. For example, any two different priors are, generically, mutually singular, and so involve incompatible world views of what is even possible. This can be expected to lead to diverging inferences from the data. The mere possibility of consistency, in the problem considered, is therefore an unexpected bonus.