properly defined intrinsic Bayes factors) unless the sample size is quite small; for very small sample sizes Berger and Pericchi (1993) recommend the 'intrinsic prior' which overcomes all these difficulties.

David Cox (Nuffield College, Oxford): In the non-nested case, the Bayesian solution appears more incisive than that based on tests, but the latter may be more informative. By taking the two models in turn as the null hypothesis, we may study whether one, both or neither model is adequate. Clearly a model could have a large Bayes factor in its favour and yet be a very bad fit.

In the non-Bayesian analysis, which has a very extensive econometric literature, the asymptotic calculations (Cox, 1961, 1962) are best replaced by simulation.

D. V. Hinkley and I (Cox and Hinkley (1978), pages 160–162) put forward a very tentative Bayesian discussion leading to subtracting from the log-likelihood ratio a penalty $\log(n/n_0)\Delta d$, where Δd is the difference in the dimensionality of the parameters involved and n_0 is a notional sample size, said very boldly to be in the range $(\frac{1}{2}, 2)$, although $(\frac{1}{2}, 5)$ might have been better. The essence of the argument was that the prior probabilities in the two models should be the same over sets of parameters giving similar predictions. I hope that in his reply Professor O'Hagan will comment.

Andrew Gelman (University of California, Berkeley) **and Xiao-Li Meng** (University of Chicago): The idea of fractional Bayes factors (FBFs) is an intriguing attempt to avoid the fundamental problem of using Bayes factors with unspecified joint densities. However, the usefulness of the Bayes factor is restricted to problems where it exists. The non-existence of the Bayes factor, as is well known, is a direct consequence of not having a density (proper or improper) defined jointly for the model indicator and the parameters within each model. The proposed solutions of this problem, therefore, have either been to complete (temporarily) such a joint specification in some way, as with partial Bayes factors, or to define different quantities that no longer have proper probability (density) interpretations, as with the FBF. The FBF is well defined in its own right but no longer has a direct Bayesian interpretation, even under a properly specified joint density (except in the limit of $b=0$). When a method slides outside the Bayesian framework it is generally found that some incoherent aspects arise. The author discusses this issue in Section 8, but we are unsure whether sequential incoherence is the only drawback (for example, Section 7.2 does not convince us that the FBF is coherent for a given sample).

From an applied point of view, we do not see the necessity of working hard to define Bayes-factor-like quantities for models without joint densities. In our experience, a full Bayesian modelling approach can always address the questions of applied interest more directly than these look-alikes. If the models being compared are nested, then we prefer conducting Bayesian inference under the larger model, using a prior distribution with preference to the region of the parameter space near the smaller model, if appropriate; an elementary illustration is given in Gelman and Meng (1994). If the models under consideration are non-nested, it is generally reasonable to expand to a larger model class with an additional continuous parameter with specific values corresponding to the original models. For instance, for the data example in the paper, Darwin's data set, we prefer the approach of Box and Tiao (1962) using the power family, which includes wide-tailed distributions, to compute the posterior distribution of the parameters of applied interest.

Of course, in model comparison problems with proper joint densities (as, for example, in discrete models in genetics), we appreciate the utility of Bayes factors in posterior inference. We are also interested in seeing methods that can address applied interests, beyond what the full Bayesian modelling approach provides, in situations with no joint densities.

Rob Kass and Larry Wasserman (Carnegie Mellon University, Pittsburgh): In his examples Professor O'Hagan takes $b=n^{-1}$. This suggests that he may find it appropriate to take the amount of information in the prior to be about the same as that in one observation. Using this heuristic in a different way leads to an interesting result, at least for nested models where, say, $\theta_1=\beta$ and $\theta_2=(\beta, \psi)$ with the first model corresponding to H_0: $\psi=\psi_0$. We transform β so that the Fisher information matrix is block diagonal when $\psi=\psi_0$ (which is always possible) and take the marginal priors on β to be equal under the null and alternative hypotheses with β and ψ independent under the alternative. Then, taking the prior on ψ to be normal centred at ψ_0 and setting the determinant of the precision matrix equal to the determinant of the Fisher information matrix for ψ (so that 'the amount of information in the prior equals the amount of information in one observation') we find that the logarithm of the Bayes factor may be approximated by the Schwarz criterion with an error of order $O(n^{-1/2})$, rather than the usual error of order $O(1)$ (Kass and Wasserman, 1992). This result suggests that the Schwarz criterion should