So without convincing guidelines for the choice of $b$ that are not based on asymptopia, I remain sceptical and tend to think that I have seen yet another ad hoc statistical tool that throws little light on the important issue of model comparison.

David Draper (University of Bath): I would like to raise the issue of why we would wish to do model comparison in the first place, because this bears on how we should do it. As Professor O'Hagan notes, Bayes factors are central to the Bayesian approach to the comparison of a finite number of models, indexed discretely. Sometimes when the purpose of the underlying investigation is inference, the vector of implied posterior probabilities for the models, given a particular choice of prior probabilities, is sufficient to answer the scientific question at hand, for example when each model corresponds to a distinct substantive theory and the goal is to summarize the weight of evidence in favour of each theory. But often the role of model comparison is more technical, leading in routine current practice to a single specification choice such as the form of the error distribution in a generalized linear model. As noted in Draper (1995) and elsewhere, an alternative—arguably preferable in many cases—would be to deal with this sort of uncertainty more smoothly by indexing the models under consideration with one or more continuous parameters and adding a layer to the modelling hierarchy corresponding to the specification uncertainty. In effect one then computes an infinite number of Bayes factors, which give rise to mixing weights used in the calculation of a weighted average posterior distribution for the quantity of direct interest. I would be interested in any comments that Professor O'Hagan might wish to make on the extent to which his concerns with discrete Bayes factors, regarding the fact that the dependence on the prior distribution does not disappear with sample size, carry over to the continuous hierarchical case in which model uncertainty is dealt with more smoothly. I was surprised by these concerns even in the discrete case, given the $O(1)$ nature of the contribution to the Bayes factor of the prior distributions on the parameters specific to each model under comparison (see equation (5)).

A. P. Dawid (University College London): Suppose that we assign a single ‘distribution’, i.e. $\sigma$-finite measure, over the full parameter space, i.e. the disjoint union of model-specific parameter spaces. If this is improper then there is no sensible way of defining marginalization and conditioning, so we will not have well-defined prior model probabilities, or Bayes factors. However, even in this case the full posterior will typically be proper, so that we will have well-defined posterior model probabilities, thus answering our real need. The specification of such a prior still requires a ‘ratio of constants’, but at least we can now see why, which may give some guidance—unlike the situation considered by O'Hagan, with its inherent arbitrariness.

Using an improper prior or an arbitrary proper prior, as above, a ‘principle of precise measurement’ applies. The full posterior, given enough data, is insensitive to the prior: it concentrates on the appropriate model, with the usual asymptotic normal form. The effect of the prior is of additive order $O(1)$ on a log-posterior-model-odds, or log-Bayes-factor when it is defined, of order $O(n)$ or $O(-\log n)$. For extensive data, this is not worth worrying about. For small data sets it is surely appropriate that prior opinion is relevant and deserves careful elicitation. It seems a retrograde step to attempt to mask an $O(1)$ effect by introducing a fractional Bayes factor with unspecified $b$, variations in which will have an effect of at least the same order—or even much larger under suggestions (b) and (c) of Section 6.

Granted that the specification of priors is still an unfamiliar and delicate task, the following approach to coherence across models may be helpful. First, specify a proper prior for the most complex model considered, or for a new model generalizing all those considered. Then specify a proper prior within each of the other models to match, as closely as possible, the induced predictive distribution for a suitable ‘minimal sample’, varying with the model considered. This idea is related to, but distinct from, that underlying the partial Bayes factor, and is fully coherent. Essentially this approach was used by Dawid and Lauritzen (1994) for comparing different decomposable graphical models.

L. I. Pettit (Goldsmiths’ College, London): I would like to congratulate Professor O’Hagan on an interesting paper that will be very useful.

All the examples considered concern global model choice. What happens if we consider local model choice? In particular I shall investigate whether a normal sample contains a single outlier. The asymptotics of Section 1.3 do not now apply.

Suppose, under $M_0$, $x_i \sim N(\mu, \sigma^2)$ for $i = 1, \ldots, n$ and, under $M_1$, $x_i \sim N(\mu + \delta, \sigma^2)$ for $i \neq j$, $x_j \sim N(\mu + \delta, \sigma^2)$. For comparison with Spiegelhalter and Smith’s (1982) method I shall take the prior proportional to $(\sigma^2)^{-(n + 1)/2}$ which leads to $r_1$ in equation (19) being constant. Note that this prior is necessary for