The main point of disagreement is with the answer to question 5, about the role of the axioms of personalistic probability. I do not at all agree that these show that one must follow the stated route. It can be very interesting to explore the consequences of simple axioms but there comes a point where what is revealed throws more light on the axioms than on the final conclusions. In particular, in the present case, the assumption that all probabilities are comparable is central. This may be reasonable in some or perhaps many contexts, but certainly not in all. As soon as it is admitted that there are different kinds of uncertainty, the argument breaks down. The arguments connected with coherent decision theory, moreover, depend upon the idea that the simple games contemplated in that theory are a reasonable representation of real-life problems, whether in science, technology or in public affairs. These are in many cases but highly idealized models of what is involved. Professor Bernardo's second line of argument is more compelling but even here seems to be tied to problems with independent and identically distributed structure and rather few of the applications I come across are of this kind. This is not at all to dismiss the interest of these arguments, rather to stress that they are considerably less compelling than Professor Bernardo's answer suggests.

My general attitude to these issues is eclectic; if more or less the same answer can be obtained from a number of different approaches this is reassuring. If very different answers are obtained, then clarification for the reasons for the differences can be very enlightening. From this point of view, one of the values of work on reference priors hinges on clarifying the link between the Bayesian and conditional confidence interval approach as summarized in the answer to question 20.

I think it would be very helpful if Professor Bernardo amplifies his answer to question 29. The assumption that information is measured in the way stated seems crucial, and from some points of view these definitions are connected with asymptotic theory; clearly Professor Bernardo has one of the stronger interpretations in mind, but is it really reasonable to suppose that the amount of information in a distribution can be captured in one number? Finally, do reference priors throw any light on the data-dependent priors used by G.E.P. Box and me in our work on transformations?

Comments on "Non-informative priors do not exist"

A.P. Dawid

Department of Statistical Science, University College, London

There is no doubt that 'reference', and other 'non-subjective', priors have played an important rôle in motivating young researchers to take a stronger interest in the Bayesian approach to Statistics. The idea of an 'objective' analysis, which somehow lets the data speak for themselves, and allows some approchement between Bayesian and non-Bayesian answers, has long had great appeal. From the pragmatic viewpoint,
it is clearly impossible to introspect deeply about every routine problem one meets, and so I sympathize with the desire to have a 'default' prior specification incorporated into Bayesian software, for example. If this idea could be put on a proper theoretical foundation, so much the better. I must therefore admit to great personal disappointment that (notwithstanding the strenuous and admirable efforts of such pioneers as Jeffreys and Bernardo) it has become clear over the years that this ideal is unattainable: no theory which incorporates non-subjective priors can truly be called Bayesian, and no amount of wishful thinking can alter this reality.

Professor José Bernardo claims to be a strong believer in foundational arguments, so I am surprised he is satisfied with a methodology which is at odds with them. For example, any way of specifying a default prior which depends (as all do) on the model under consideration must violate the likelihood principle (see e.g. Berger and Wolpert, 1984), a mainstay of Bayesianism. Professor Bernardo's own reference priors depend further on the parameter function being considered. This means that the associated 'posteriors' do not even obey the normal rules of probability. For the model $X_i \sim N(\mu_i, 1)$ independently, the reference posterior distribution for $\theta = \sum \mu_i^2$ cannot be found by marginalization from the joint reference posterior for the vector $\boldsymbol{\mu} = (\mu_i)$. Should calculation of this margin be outlawed? If so, why? How should we calculate the posterior probability of an event of the form "$\mu \in A$", when $A = \{ \mu: \sum \mu_i^2 \leq k \}$? We get different answers depending on whether we use the reference prior for $\mu$ or that for $\sum \mu_i^2$ (and, presumably, yet another answer if we use that for the indicator function of $A$ – at any rate, this cannot agree with both the other answers). What if we perturb the boundary of $A$ very slightly so that it is no longer determined by the value of $\sum \mu_i^2$? Can we countenance a discontinuous jump in the probability of $A$? What is the use of a distribution for $\boldsymbol{\mu}$, anyway, if we cannot use it to assign probabilities to arbitrary events?

Professor Bernardo suggests that the 'marginalization paradox' (MP) can always be avoided by the use of his reference prior for the parameter of interest. This has not been demonstrated, in general. Section 3 of Dawid et al. (1973) exhibits cases where, for suitable functions $z$ of the data and $\zeta$ of the parameter, with the density $p(z|\zeta)$ of $z$ depending only on $\zeta$, priors may be found for which the marginal posterior for $\zeta$ depends only on $z$, but for no such prior can this posterior have $p(z|\zeta)$ as a factor, which would be required to evade MP. It is still possible to evade MP by using an arbitrary proper prior, but this operates by a different mechanism: it is then impossible that the marginal posterior of $\zeta$ depend on $z$ alone, so the possibility of MP does not even arise. The reference analysis would have to behave similarly if Professor Bernardo's conjecture is valid. Would this behaviour be acceptable?

Professor Bernardo makes several references to the idea of an improper prior as an approximation to a proper one. However, such arguments are delicate and prone to fatal pitfalls: see, e.g., Stone (1982). A recent account of logical issues and difficulties associated with MP and approximability may be found in Dawid et al. (1996).

In summary, the idea of a 'default prior' is here to stay, for very good pragmatic reasons; but there can never be a fully consistent theory of such priors, so they must be treated with great caution. I fully agree with the title of the article under discussion.
**Non-informative priors do not exist – discussion of a discussion**

J.K. Ghosh

*Indian Statistical Institute and Purdue University*

What a honest but witty and civilized discussion of topics that many of us regards as important and most of us concede as controversial. Professor Bernardo, as Socrates, may not have convinced his sceptical pupils completely but he has certainly put up a strong case for doing what goes under the name of default or automatic or non-subjective Bayesian analysis. Few of us would disagree either with his description of its limited aims of supplementing rather than replacing subjective Bayesian analysis or his concern with useful non-subjective posteriors instead of noninformative priors. However, difficulties remain even if one gives up noninformative priors in favour of data-dominated posteriors. To keep my discussion simple, I will focus on this and certain technical questions arising from this. I will then discuss briefly the question of truncating the improper priors.

By data-dominant posteriors Professor Bernardo clearly means the posteriors that arise by maximizing the Lindley information measure which is an average of the Kullback–Leibler divergence of the prior and the posterior. This procedure was introduced in Professor Bernardo's seminal paper of 1979 and a somewhat modified and very clear algorithm was provided in Berger and Bernardo (1989, 1992b). It is true that the proposed criterion is elegant and general and all the applications have produced posteriors that seem to be satisfactory. Unfortunately, we still donot understand fully why this should be so. Let me explore briefly internal validation through coherence, avoidance of the marginalization paradox and proper posterior and external validation through frequentist considerations over repetitions as in Berger and Bernardo (1989).

In problems which are invariant under an amenable group of transformations, it is known that inference based on the right-invariant Haar measure as the prior is both