ELSEVIER

Journal of Statistical Planning and Inference 65 (1997) 159-189

journal of statistical planning and inference

Non-informative priors do not exist A dialogue with José M. Bernardo[1]

Preamble

Professor José Bernardo, who is the originator and. with James Berger, a driving force behind the use of what are known as 'reference priors', shared a ride with Telba Irony and Nozer Singapurwalla from Pittsburgh, PA. to Washington, DC, sometime during the latter part of 1995. After the usual gossip, which statisticians of all persuasions tend to relish, the discussion turned to the topic of 'dishonest priors', and their increasing encroachment on the Empire of Chance. The discussion evolved in a Socratic tradition, with José playing the role of Socrates (albeit answering the questions in this particular dialog), and his driving partners, the pupils; fortunately José was not driving. This question and answer format of discussion turned to be very fruitful, and even more enjoyable than gossip, because José's knowledge of the topic and its historic evolution, not to mention his passion for statistics (among other things), provided a useful perspective on a topic of much controversy. José was then asked to write up his perspective as a contribution for the Discussion Forum of JSPI, to which his suggestion was that we reproduce the dialogue and write it up as such. We thought that this suggestion was a great idea and so, when José visited Washington DC in May 1996, the dialogue was reconstructed. Given below is an account of this reconstruction; we feel that those of us who are not cognoscenti about dishonest priors may benefit from this conversational overview.

Telba Z. Irony
Nozer D. Singpurwalla

Question 1. It is often said that *non-informative* priors do not exist and yet under this label, plus many others, such as *conventional, default, flat, formal, neutral, non-subjective, objective* and *reference* priors, they do get used. Does this imply that the users of non-informative priors are really not honest Bayesians?

Answer. I would not say that they are dishonest but, too often, they are not precisely aware of the implications of their use. By definition, 'non-subjective' prior distributions,

[1] José M. Bernardo is Professor of Statistics at the University of Valencia. Research partially funded with grant PB93-1204 of the DGICYT, Madrid, Spain. The author gratefully acknowledges stimulating comments by Jim Berger, Gauri Datta, Phil Dawid, Dennis Lindley, Angelika van der Linde and Elias Moreno to an earlier version of this manuscript.

0378-3758/97/$17.00 © 1997 Elsevier Science B.V. All rights reserved
PII S0378-3758(97)00074-8

are *not* intended to describe personal beliefs, and in most cases, they are *not even proper* probability distributions in that they often do not integrate one. Technically, they are *only* positive functions to be formally used in Bayes theorem to obtain 'non-subjective posteriors' which, for a *given model*, are supposed to describe whatever the data 'have to say' about some *particular quantity of interest*, either a function of the parameters in the model, or a function of future observations. Whether or not they achieve this goal or, indeed, whether or not this goal is at all achievable, is often a matter of debate.

Question 2. Are non-subjective posteriors always proper?

Answer. They *should* be! Indeed, a proper posterior for any minimum-size sample should be the first property required from any method of deriving non-subjective priors. But you should realize that a naïve use of 'standard' non-subjective priors *may* lead to improper posteriors. Casella (1996) describes how, in a components of variance problem, this has lead to unaware users of Gibbs sampling to obtaining nice pictures of posterior distributions that do not exist! Actually, Dennis Lindley told me that he knew of this components of variance example since the 1950s, and it is mentioned in Berger (1985, p. 187), but people seem to take for granted a crucial property – the propriety of the posterior – which *must* however be carefully checked whenever an improper prior is used.

Question 3. Are there any non-subjective *priors* that are proper?

Answer. There are indeed. The simplest example is the Beta distribution Be $(\theta|\frac{1}{2},\frac{1}{2})$, widely regarded as the appropriate non-subjective prior to make inferences about the parameter θ of a binomial model. In fact, proper non-subjective priors are usually found whenever the parameter space is bounded, although this is not a general rule; the non-subjective prior typically recommended for the parameter θ of a negative binomial model is $\pi(\theta) = \theta^{-1}(1-\theta)^{-1/2}$, which is improper although the parameter space is bounded, while a sensible non-subjective prior to make inferences about the ratio of two multinomial parameters, turns out to be proper even though the parameter space $]0, \infty[$ is not bounded. Actually, one could always work with *proper* non-subjective priors if the parameter spaces were taken to be appropriately chosen bounded sets. For example, the standard non-subjective prior for a real location parameter is uniform on $\Re$, which is improper; however, if given some experimental measures, inferences are made on the true value of some physical quantity a more realistic parameter space would be $[0, c]$, for some context-dependent constant c, and the presumably appropriate non-subjective prior, a uniform on $[0, c]$, would indeed be proper. Similarly, in the negative binomial setting where the probability of success must be strictly positive, a parameter space of the form $[\varepsilon, 1]$, for some $\varepsilon > 0$ would lead to a proper non-subjective prior.

Question 4. How do you interpret probability?

Answer. Naturally, as a measure of personal belief. Of course, this does not mean that I would systematically be prepared to bet in terms of non-subjective posterior distributions, because my *personal* beliefs may well be not closely approximated by any particular non-subjective prior.

Question 5. I agree that betting quotients may not reflect true belief, but do you subscribe to the axioms of probability?

Answer. Yes, I do. I am a strong believer in foundational arguments: the intellectual strength of the Bayesian argument comes directly from the fact that mathematical logic requires one to express *all* uncertainties by means of *probability* measures. In fact, there are two independent arguments to support this claim:

(i) *Coherent decision theory*: If you try to guarantee that your decision making criteria are sensible - in that they meet some intuitively appealing axioms or if you want to avoid 'inadmissible' decisions, then you *must* express all your uncertainties by probabilities. For specific details, you can look at Savage (1954), Fishburn (1981, 1986), or Bernardo and Smith (1994, Ch. 2), and references therein.

(ii) *Representation theorems*: If you accept a probabilistic description of the behaviour of observables – as all statisticians presumably do – and you want to describe mathematically the idea that some observations are 'similar' in some sense and hence some type of prediction is possible then the general representation theorem tells you that these 'exchangeable' observations are a random sample from some underlying model, indexed by some parameter defined as the limit of some function of the observations, and *there exists a prior* probability distribution over such a parameter. Key references are de Finetti (1937), Hewitt and Savage (1955), Smith (1981) and Diaconis (1988); you can get an overview from Bernardo and Smith (1994, Ch. 4), and references therein.

These are proven *existence results*; they imply that the common sentence 'a prior does not exist' is a mathematical *fallacy*: for mathematical consistency one *must* be a Bayesian. However, these are *only* existence results; they leave open the question of *specifying* a particular prior in each problem.

Question 6. Don't the axioms imply that the probability of a tautology should be 1 and, thus, that priors have to be proper?

Answer. The natural axioms do not: they only lead to finite additivity, which is compatible with improper measures; however, the further sensible assumption of *conglomerability* leads to σ-additivity and, hence, to proper measures; some signposts to this debate are Renyi (1962), Heath and Sudderth (1978, 1989), Hartigan (1983), Cifarelli and Regazzini (1987), Consonni and Veronese (1989) and Lindley (1996). Nevertheless, it must be stressed that what really matters is the posterior of the quantity of interest – which under σ-additivity must certainly be proper – , because this is what you have to use either in inference or in decision making.

Question 7. But those proper non-subjective posteriors are often derived from improper priors; how should one interpret the improper priors?

Answer. One should *not* interpret *any* non-subjective prior as a probability distribution. Non-subjective priors are merely positive fractions which serve to produce non-subjective posteriors by formal use of Bayes theorem, and 'sensible' non-subjective posteriors are *always* proper.

Question 8. What do you mean by a 'sensible' non-subjective posterior?

Answer. One that – after careful scrutiny of its properties – you would be prepared to use for scientific communication. Of course, there is no way to give a formal definition for this; indeed, an important part of the discussion on methods for deriving non-subjective priors is based on the analysis of the statistical properties of the posteriors they produce in specific, 'test case' examples.

Question 9. If the prior to posterior conversion describes the change in one's betting behaviour (and this is a debatable issue), should that prior also not be proper?

Answer. Non-subjective priors are *limits.* Any sensible non-subjective prior may be seen as some appropriately defined limit of a sequence of proper priors. In fact, as I have mentioned before, they are only improper because of the use of convenient, typically unbounded parameter spaces. If we tried to be more realistic, and worked with appropriately chosen bounded parameter spaces, non-subjective priors would always be proper; if one prefers to work with conventional parameter spaces, increasing sequences of bounded approximations to those parameter spaces may be used to provide sequences of *proper* priors which converge to the corresponding improper non-subjective priors in the precise sense that, for any data set, the sequence of posteriors thus produced converges to the corresponding non-subjective posterior.

Question 10. Why then have improper priors entered our lives? Can we not do away with them by concentrating only on compact spaces?

Answer. We could indeed, but this would probably be mathematically inconvenient. But, again, the main question is *not* whether non-subjective priors are proper or improper, but whether or not they lead to sensible non-subjective posteriors. Actually, if an improper prior leads to a posterior with undesirable properties, the posterior which would result from a proper approximation to that prior (like that obtained by truncation), will typically have the same undesirable properties: for instance, the posterior of the sum of the squares of normal means $\phi = \sum \mu_i^2$ based on a joint uniform prior on the means $\pi(\mu_1, \dots, \mu_k) \propto 1$ is extremely unsatisfactory as a non-subjective posterior for ϕ (Stein, 1959), but so is that based on the *proper* multinormal prior $\pi(\mu_1, \dots, \mu_k) \propto \prod_i \mathrm{N}(\mu_i \,|\, 0, \sigma)$, for large σ. Proper or improper, we need non-subjective priors which appropriately represent a lack of relevant prior knowledge about the quantity of interest *relative* to that provided by the data.

Question 11. What do you mean by a prior describing a lack of knowledge?

Answer. The contribution of the data in constructing the posterior of interest should be 'dominant'. Note that this does *not* mean that a non-subjective prior is a mathematical description of 'ignorance'. *Any* prior reflects some form of knowledge. Non-subjective priors try to make precise the type of prior knowledge which, for a *given model* and for a *particular inference problem* within this model, would make data dominant. They may also be seen as a contribution to robust Bayesian methods (Berger, 1994).

Question 12. We now see your point: namely, that non-subjective priors are those for which the contribution of the data are posterior dominant for the quantity of interest. This is sensible, but to construct non-subjective priors does one need to consider data?

Answer. Non-subjective priors do not typically depend on the data, but on the probabilistic model which is assumed to have produced them: as one would expect, the prior which makes the data posterior dominant for the quantity of interest usually depends both on the model assumed for the data and on the quantity of interest.

Question 13. It appears that, in using the prior as a mere technical device in a formal use of Bayes' law, and with the aim of making the posterior data dominant, one is proposing a paradigm for inference which is not in the spirit of Bayes (and Laplace) nor is it in the spirit of Fisher, or Neyman–Pearson–Wald. If so, may one conclude that a non-subjective Bayesian is a sensible *pseudo-Bayesian*?

Answer. No, I think he is totally Bayesian. Non-subjective Bayesian analysis is just a part – an important part, I believe – of a healthy *sensitivity analysis* to the prior choice: it provides an answer to a very important question in scientific communication, namely, what could one conclude from the data *if* prior beliefs were such that the posterior distribution of the quantity of interest were dominated by the data.

Question 14. What are the historical precedents of non-subjective priors?

Answer. The pioneers were Bayes (1763), who considered a uniform prior in a binomial setting, and Laplace (1812), who used an improper uniform prior in the normal case. Nobody considered using priors that were different from the uniform in those days when, by the way, a large part of statistical inference was based on inverse probability calculations.

Question 15. So Laplace did use uniform priors that were improper?

Answer. Yes, although he seemed to be aware that this was only a sensible approximation to using a proper uniform prior in the *bounded* parameter space which would more closely reflect the underlying physical problem.

Question 16. How does this connect with Laplace's 'principle of insufficient reason'?

Answer. The 'rationale' for using a uniform prior was that any other prior would reflect specific knowledge. Of course, as I stated before, any prior reflects *some* knowledge; what happens here is that in any *location* problem, the uniform prior is precisely that which makes the data posterior dominant.

Question 17. When did problems with uniform priors surface?

Answer. By the early 1920s it was widely realized that the universal use of a uniform prior did not make sense. Since most statisticians were not prepared to use personal priors in scientific work, alternative 'objective' methods of statistical inference were produced, which only depended on the assumed probability model; this gave rise to both fiducial and frequentist inference. It was not until the 1940s that Jeffreys (1946) produced an alternative to using the uniform as a non-subjective prior; however, Jeffreys was a physicist, barely integrated in the world of academic statistics, and his work did not achieve the impact is deserved.

Question 18. What did Jeffreys do?

Answer. He was motivated by invariance requirements and suggested, using differential geometry arguments, a solution which provides a non-subjective prior known as Jeffreys' prior. He then proceeded to a detailed investigation of the consequences of such an approach (Jeffreys, 1961).

Question 19. What are these invariance requirements? Are they really crucial?

Answer. They are invariance under one-to-one transformations, and invariance under sufficient statistics. I certainly believe that those properties are crucial for 'sensible' non-subjective distributions, to the point that one should not seriously consider a proposal for non-subjective Bayesian inference which does not satisfy them:

(i) *Invariance under one-to-one transformations.* If $\theta = \theta(\phi)$ is a one-to-one function of ϕ, then $\pi(\phi|x)$ is *logically equivalent* to $\pi(\theta|x) = \pi(\phi|x)|d\phi/d\theta|$; hence, the non-subjective posterior for ϕ directly obtained from $p(x|\phi)$ and the non-subjective posterior for θ obtained from the same model reparametrized in terms of θ *must* be related by that equation.

(ii) *Invariance under sufficient statistics.* If t is *sufficient* statistic for the model $\pi(\phi|x)$, then the non-subjective posterior $\pi(\phi|x)$ obtained from model $p(x|\phi)$ *must* be the same as the non-subjective posterior $\pi(\phi|t)$ obtained from $p(t|\phi)$.

Some pointers to the literature on the role of invariance in the selection of non-subjective priors are Hartigan (1964), Jaynes (1968), Dawid (1983) and Yang (1995).

Question 20. Can you give an intuitive explanation of Jeffreys' prior?

Answer. His invariance requirements are easy to understand, but he did not offer an intutively convincing explanation for his particular choice; however, a modern description by Kass (1989) offers a heuristic explanation based on the idea that 'natural' volume elements, defined in terms of Fisher's matrix, should have equal prior probability. Moreover, when applicable (continuity and appropriate regularity conditions), the one-dimensional version of Jeffrey's prior has been justified from many different viewpoints: these include Perks (1947), Lindley (1961), Welch and Peers (1963), Hartigan (1965), Good (1969), Kashyap (1971), Box and Tiao (1973, Section 1.3), Bernardo (1979), Kass (1990), Wasserman (1991) and Clarke and Barron (1994). The derivation of the one-dimensional Jeffreys' prior by Welch and Peers (1963), as that prior for which the coverage probabilities of one-sided posterior credible intervals are asymptotically as close as possible to their posterior probabilities, may be specially appealing to frequentist trained statisticians: this means that under Jeffreys' one parameter prior, and for large sample sizes, an interval with posterior probability $1 - \alpha$ may be *approximately* be interpreted as a confidence interval, in the Neyman–Pearson sense, with significant level α.

Question 21. You have just mentioned that, from a technical point of view, Jeffreys prior is related to Fisher's information matrix. Is Fishers' information matrix related to the notion of 'lack of information'?

Answer. Not directly. The connection comes from the role of Fisher's matrix in asymptotics. By the way, this should be simply called 'Fisher's matrix', not 'Fisher's information matrix': it is only directly related to general information measures under normal assumptions or in regular asymptotic conditions.

Question 22. Is Jeffreys' prior a flat prior?

Answer. I guess you mean 'nearly uniform'. Then no, generally it is not, unless the parameter is a location parameter. I will use your question to stress that 'flat' priors – typically uniform or log-uniform – which are too often used as being synonymous with 'non-informative' priors, may be *very* informative on non-location parameters. For instance, as I have mentioned before, a 'flat' prior on the means of a multivariate normal implies strong knowledge about their sum of squares, thus producing Stein's (1959) paradox.

Question 23. Then in what sense is Jeffreys' prior neutral, i.e., reflects a lack of knowledge?

Answer. It does *not* reflect 'lack of knowledge', but it may be argued that – with one parameter and under regularity conditions – Jeffreys' prior describes the type of prior knowledge which would make the data as posterior-dominant as possible. Thus, the corresponding posterior distribution may be argued to provide a benchmark, a 'reference', for the class of the posterior distributions which may be obtained from other, possibly subjective, priors.

Question 24. What is the problem with Jeffreys' prior and in which cases is it appropriate?

Answer. Jeffreys himself realized that his proposal only works, and then under regularity conditions, in one-parameter continuous problems. He suggested a collection of ad hoc rules to deal with multiparameter problems, with mixed results. Moreover, he seemed to be convinced that a *unique* appropriate non-subjective prior could be defined for any given model, whatever the quantity of interest. This was later seen to be not true.

Question 25. What are the developments after Jeffreys'?

Answer. For a while, it was thought that clever analysis of multiparameter problems using some combination of Jeffreys' proposals would produce appropriate non-subjective priors for any problem. Lindley's (1965) book was an explicit attempt is this sense, and it *does* prove that most standard 'textbook' inference problems have a non-subjective Bayesian solution within his framework, and one which produces credible intervals which are often *numerically* either identical or very close to their frequentist counterparts. But then, in the early 1970s, the marginalization paradoxes emerged.

Question 26. What are the marginalization paradoxes?

Answer. They may collectively be seen as a proof that the original idea of a *unique* non-subjective prior for each model is untenable: we may only agree on a unique non-subjective prior for each quantity of interest within a model. A simple example of marginalization paradox is provided by the standardized mean $\phi = \mu/\sigma$ of a normal distribution: Stone and Dawid (1972) showed that the posterior distribution of ϕ only depends on the data through some statistics t, whose sampling distribution only depends on ϕ. Hence, one would expect that inferences derived from the model $p(t\,|\,\phi)$ would match those derived from the full model $\mathrm{N}(x\,|\,\mu,\sigma)$, but they proved that this is not possible if one uses the 'standard' non-subjective prior $\pi(\mu,\sigma) = 1/\sigma$, which everybody agrees is the appropriate non-subjective prior to make inferences about either μ or σ. It was immediately seen (Dawid et al. 1973) that marginalization paradoxes are ubiquitous in multiparameter problems: any future development of non-subjective Bayesian analysis would have to come to terms with them.

Question 27. Was your own work on *reference* distributions a reaction to this?

Answer. It was not a direct reaction, but I was certainly influenced by these results. In the mid 1970s, as a part of my Ph.D. work on Bayesian design of experiments, I become interested in non-subjective 'non-informative' priors. The marginalization paradoxes made obvious to me that some new work in that area was necessary: reference analysis was the result.

Question 28. Can you explain what do you mean by reference analysis?

Answer. Reference analysis may be described as a method to derive model-based, non-subjective *posteriors*, based on information-theoretical ideas, and intended to describe the inferential content of the data for scientific communication. It is, to the best of my knowledge, the only general method available which has the required invariance properties and successfully deals with the marginalization paradoxes.

Question 29. This is a very convincing statement in favour of your paradigm, but what do you mean by the inferential content of the data? How many this be quantified?

Answer. In the sense, each possible answer to that pair of related questions may be the basis for a method to derive non-subjective posteriors. I personally believe that the inferential content of the data is appropriately measured by the *amount of information* they provide on the quantity of interest, where the word 'information' is used in the technical sense of Shannon (1948) and Lindley (1956). This was the starting point for the definition of a *reference* posterior (Bernardo, 1979).

Question 30. Can you be more explicit on the relationship between reference distributions and information-theoretical ideas?

Answer. The amount of information to be *expected* from the data is naturally a function of the prior knowledge, as described by the prior distribution: the more prior information available, the less information may one expect from the data. With only one real-valued parameter, one may unambiguously define a limit functional which measures, in terms of the prior distribution, the amount of *missing information* about the parameter which data from a given model could possibly be expected to provide; the reference prior is that which maximizes the missing information. The multiparameter case is handled by recursively using the one-parameter solution.

Question 31. How do reference priors differ from other proposals? In particular, how do they differ from Jeffreys' priors?

Answer. The reference prior approach is totally general and, as far as I am aware, it includes within a single framework all generally accepted non-subjective solutions to specific cases. In one-parameter problems, the reference prior reduces to Jaynes (1968) *maximum entropy* prior if the parameter space has a finite number of points, and it reduces Jeffreys' prior in the continuous regular case. In regular continuous multiparameter problems, one often obtains the solutions which Jeffreys suggested using ad hoc arguments rather than his general multivariate rule. Moreover, reference analysis can deal with non-regular cases which cause problems for other methods.

Question 32. Can you give some examples of reference priors in the one-parameter continuous case?

Answer. As I have just mentioned, under regularity conditions to guarantee asymptotic normality, the reference prior is simply Jeffreys' prior, namely

$$\pi(\theta) \propto \left(\boldsymbol{E}_{x|\theta}\left[-\frac{\mathrm{d}^2}{\mathrm{d}\theta}\log p(x|\theta) \right] \right)^{1/2},$$

but I will give you a couple of non-regular examples:

(i) *Uniform distribution on* $[\theta - a, \theta + a]$. The reference prior is then uniform on $\Re$, and the reference posterior is uniform over the set of θ values which remain feasible after the data have been observed (Bernardo and Smith, 1994, p. 311).

(ii) *Uniform distribution on* $[0, \theta]$. The reference prior is then $\pi(\theta) \propto \theta^{-1}$, and the reference posterior is a Pareto distribution (Bernardo and Smith, 1994, p. 438).

Question 33. You have sketched the derivation of reference posteriors associated to models with only one real-valued parameter, and stated that those are invariant under reparametrization, but how do you deal with nuisance parameters?

Answer (Recursively). The idea is very simple, although there are delicate technical issues involved. Consider the simplest case; suppose that you are interested in the reference posterior distribution $\pi(\phi|x_1, \ldots, x_n)$ of some quantity ϕ given a random sample from a model $p(x|\phi, \lambda)$, which contains one real-valued nuisance parameter $\lambda \in \Lambda \subset \Re$. Working conditionally on ϕ, this is a one-parameter problem, and hence the one-parameter solution may be used to provide a *conditional* reference prior $\pi(\lambda|\phi)$. If this is proper, then it may be used to integrate out the nuisance parameter λ and obtain a model with one real-valued parameter $p(x|\phi)$ to which the one-parameter solution is applied again to derive the *marginal* reference prior $\pi(\phi)$; the desired reference posterior is then simply

$$\pi(\phi|x_1, \ldots, x_n) \propto \pi(\phi) \int_\Lambda \textstyle\prod_{i=1}^{n} \{p(x_i|\phi, \lambda)\} \pi(\lambda|\phi)\, \mathrm{d}\lambda.$$

If $\pi(\lambda|\phi)$ is not proper, the procedure is performed within an increasing sequence of bounded approximations $\{\Lambda_j, j = 1, 2, \ldots\}$ to the nuisance parameter space Λ, chosen such that $\pi(\lambda|\phi)$ is integrable within each of them; the reference posterior is then the limit of the resulting sequence $\{\pi_j(\phi|x_1, \ldots, x_n), j = 1, 2, \ldots\}$ of posterior distributions (Berger and Bernardo, 1989, 1992b).

Question 34. Does this mean that, within a *single* model, you may have as many reference priors as possible parameters of interest?

Answer. It does indeed, Given a model, say $p(x|\theta_1, \theta_2)$, the reference algorithm provides a reference *posterior* distribution for *each* parameter of interest $\phi = \phi(\theta_1, \theta_2)$, and those may well correspond to different priors, because beliefs which maximize the missing information about $\phi = \phi(\theta_1, \theta_2)$ will generally differ from those which maximize the missing information about $\eta = \eta(\theta_1, \theta_2)$, unless η happens to be a one-to-one function of ϕ.

Note also that, as I mentioned before, using different priors for different parameters of interest is the *only* way to have non-subjective priors which avoid the marginalization paradoxes. For instance, in a normal model, $\mathrm{N}(x_i|\mu, \sigma)$, the reference posterior for μ is the Student distribution $\mathrm{St}(\mu|\bar{x}, (n-1)^{-1/2}s, n-1)$, obtained from the 'conventional' improper prior $\pi(\sigma|\mu)\pi(\mu) = \sigma^{-1}$, while the reference posterior for $\phi = \mu/\sigma$ is obtained from $\pi(\sigma|\phi)\pi(\phi) = (2 + \phi^2)^{-1/2}\sigma^{-1}$, a *different* improper prior, producing a reference posterior for ϕ which *avoids* the marginalization paradox that you would get if you used again the conventional prior (Bernardo, 1979).

Question 35. We now see how to deal with a single nuisance parameter, but how do you proceed when there are more than one?

Answer. The algorithm I have just described may easily be extended to any number $\{\lambda_1, \dots, \lambda_m\}$ of *ordered* nuisance parameters: get the one-parameter conditional reference prior $\pi(\lambda_m | \phi, \lambda_1, \dots, \lambda_{m-1})$ and use this to integrate out λ_m; get $\pi(\lambda_{m-1} | \phi, \lambda_1, \dots, \lambda_{m-2})$ and use this to integrate out λ_{m-1}; continue until you get $\pi(\phi)$; then use

$$\pi(\lambda_m | \phi, \lambda_1, \dots, \lambda_{m-1}) \pi(\lambda_{m-1} | \phi, \lambda_1, \dots, \lambda_{m-2}) \times \cdots \times \pi(\lambda_1 | \phi) \pi(\phi)$$

in Bayes theorem, and marginalize to obtain the desired reference posterior $\pi(\phi(x_1, \dots, x_n)$.

The result *might* possibly depend on the *order* in which the nuisance parameters are considered which, in that case, should reflect their order of importance in the problem analysed, the least important being integrated out first. We have found however that this is usually *not* the case: in most problems, the reference posterior of the quantity of interest is independent of the order in which the nuisance parameters are considered.

Question 36. Can you give some examples of this?

Answer. In a multinomial model, $\mathrm{Mu}(r_1, \dots, r_m | \theta_1, \dots, \theta_m, n)$, the reference posterior for, say, θ_1, is the Beta distribution $\mathrm{Be}(\theta_1 | r_1 + \frac{1}{2}, n - r_1 + \frac{1}{2})$ and this is independent of the order in which the other θ_i's are considered (Berger and Bernardo, 1992a). Note, by the way, that this does *not* depend on the irrelevant number of categories m as the posterior from Jeffreys' multivariate prior does; thus, the reference algorithm avoids this type of *agglomeration paradox* typically present in other proposals. Similarly, within the same model, the reference posterior for $\phi = \theta_1/\theta_2$ is the Beta distribution of the second kind

$$\pi(\phi | r_1, \dots, r_m, n) \propto \frac{\phi^{r_1 - 1/2}}{(1 + \phi)^{r_1 + r_2 + 1}},$$

(which, again, does not depend on m, but corresponds to a *different* prior), and this is independent of the order in which the nuisance parameters are considered (Bernardo and Ramón, 1996). Many more examples are referenced in Yang and Berger (1996).

Question 37. What has now happened to the invariance properties on which you insisted before?

Answer. They are still there. The reference posterior of any quantity of interest ϕ does not depend on whether one uses the full model or the joint sampling distribution of a set of sufficient statistics. Moreover, for any model $p(x | \phi, \lambda_1, \dots, \lambda_m)$, the reference posterior $\pi(\phi | x)$ does not depend on the particular parametrization chosen for each of the nuisance parameters and, besides, if $\theta = \theta(\phi)$ is a one-to-one transformation of ϕ, then $\pi(\theta|x) = \pi(\phi|x)|\,\mathrm{d}\phi/\mathrm{d}\theta|$. Datta and Ghosh (1996) have recently shown that these invariance properties are often *not* shared by other proposed methods to derive non-subjective posteriors.

Question 38. How do you compute, in practice, reference distributions?

Answer. Reference priors only depend on the model through its asymptotic behaviour; essentially, if you know the asymptotics of your model, then you may easily find its associated reference priors. Under regularity conditions for asymptotic normality, any reference prior may be obtained from a relatively simple algorithm in terms of Fisher's matrix (Berger and Bernardo, 1992b). However, the derivation of reference priors in non-regular or complex models may be a difficult mathematical problem.

Of course, once you have obtained the appropriate reference prior for some quantity of interest, you simply use Bayes theorem and marginalize to derive the required reference posterior. It turns out that, within the exponential family, reference priors often correspond to some limiting form of the corresponding natural conjugate family and, in that case, the corresponding reference posteriors may often be obtained in closed form. When this is not the case, numerical reference posteriors may be efficiently obtained using MCMC sampling–resampling techniques, as described by Stephens and Smith (1992).

Question 39. We now have a procedure to derive reference posterior distributions when no prior information is available about the parameter of interest; however, even for scientific communication, one may sometimes want to use some *partial* information (possibly intersubjectively agreed). Can reference analysis deal with this situation?

Answer. It surely can; you define the reference prior under partial information as that which maximizes the missing information subject to whatever constraints are imposed by the information assumed. If the restrictions take the form of expected values, then explicit forms for the corresponding restricted reference priors are readily obtained (Bernardo and Smith, 1994, pp. 316–320). Note that restricted reference analysis typically leads to *proper* priors; for instance, in a location model, the reference prior which corresponds to the partial information provided by the first two moments of the unknown parameter is the *normal* distribution with those moments.

Question 40. Please, talk about the new developments on reference priors.

Answer. I can see several directions in which further research is needed:

(i) *Bounded approximations*: I have mentioned before that to implement the reference algorithm in multiparameter problems when the conditional reference priors are not proper, a bounded approximation to the parameter space in which the conditional reference priors are integrable is required. It may be seen (Berger and

Bernardo, 1989) that the result *may* depend on the bounded approximation chosen. Although in a specific model it is usually clear what the 'natural' bounded approximation is – and this should be the *same* for all parameters of interest within the same model – a general definition of the appropriate bounded approximation is needed.

(ii) *Grouping*: With many parameters, one may apply the reference prior algorithm by lumping the parameters in just two groups (parameters of interest and nuisance parameters), or one may lump them into any number groups and proceed sequentially (Berger and Bernardo, 1992a–c). There is evidence to suggest, however, that one should *not* group the parameters but proceed recursively using the one-parameter solution as I have described to you before; this seems to guarantee both admissible coverage properties and the absence of marginalization paradoxes, but further research is needed to substantiate this point.

(iii) *Prediction and hierarchical models*: A reference prior is technically defined for an ordered parametrization suggested by the problem of interest. What are the appropriate ordered parametrizations to use in prediction and hierarchical model problems? Again, although some answers are available in specific cases, the general strategy is not clear.

(iv) *Model choice*: Reference distributions are not directly applicable to model choice between models of different dimensionalities: indeed, reference distributions are typically only defined up to proportionality constant, and those constants become relevant in this case. Nice results are available however (Bernardo, 1996) by posing the question as a decision problem, and working with the reference posterior of the quantity of interest implied by the corresponding utility function.

(v) *Numerical reference analysis*: The derivation of reference priors may sometimes be a difficult mathematical problem, but numerical reference posteriors may be obtained, in principle, by simulation methods. This is easily done in one or two parameters, but the general problem is not trivial (Efstathiou, 1996, Ch. 5), as computational explosion has to be avoided in higher dimensions.

Question 41. Is there anything to be said about the long-term frequentist properties of non-subjective posteriors? I realize that good Bayesians should not be raising this type of a question but, politically speaking, it may be wise to raise the issue.

Answer. Politics aside, this is a very interesting issue, and one that is central to discussions on *comparative* statistical inference. Interest on the frequentist coverage probabilities of credible intervals derived from non-subjective posteriors has a long history; key references include the pioneering work by Welch and Peers (1963) that I have already mentioned, Peers (1965). Hartigan (1966), Tibshirani (1989), Ghosh and Mukerjee (1992), Mukerjee and Dey (1963), Nicolau (1993), and Datta and Ghosh (1995). The coverage probabilities of credible regions has often been an important element in arguing among competing non-subjective posteriors, as in Berger and

Bernardo (1989) or Ye and Berger (1991) and, indeed, discussions on the coverage properties of non-subjective priors have now become common. Reference posteriors have consistently been found to have attractive coverage properties – what may be seen as a form of *calibration* – but, as far as I am aware, no general results have been established.

Question 42. Is there any other issue on comparative inference over which non-subjective priors may have a bearing?

Answer. A very important one is the procedure used to eliminate nuisance parameters. The marginalization paradox examples may be used to demonstrate that non-Bayesian methods to eliminate nuisance parameters (plug-in estimates, profile likelihood, naïve integrated likelihood and the like) are often *inconsistent within their own paradigms* in that the resulting 'marginal' likelihood may differ from the simplified likelihood. For example, the sampling distribution of the sample coefficient of correlation r in a bivariate normal model depends only on the population coefficient of correlation ρ, so that non-Bayesian statisticians would presumably consider $p(r|\rho)$ to be an 'exact' 'marginal' likelihood from which inferences about ρ could be made. Yet, the more sophisticated non-Bayesian techniques to eliminate nuisance parameters fail to derive $p(r|\rho)$ (Efron, 1993), while integration of the nuisance parameters with the conditional reference priors (taken in any order) easily produces the 'exact' marginal likelihood (Bayarri, 1981; Lindley, 1965, pp. 215–219; Bernardo and Smith, 1994, pp. 363–364).

Question 43. What about admissibility?

Answer. Non-subjective priors are sometimes criticized on the grounds that, since they are often improper, they may lead, for instance, to inadmissible estimates. We have seen, however, that sensible non-subjective priors are, in an appropriate sense, limits of proper priors; regarded as a 'baseline' for admissible inferences, non-subjective posteriors need not be themselves admissible, but only arbitrarily close to admissible posteriors. That said, admissibility is not really the relevant concept: truncating the parameter space may lead to technically admissible but very unsatisfactory posteriors, as in the sum of squares of normal means example I described before. Besides, admissibility crucially depends on the loss function; thus, if one is really interested in estimation, one should explicitly work in terms of the corresponding decision problem, with an appropriate, context dependent, loss function. To deal with those problems, reference analysis may be extended to define *reference decisions*, as those optimal under the prior which maximizes the *missing utility* (Bernardo, 1981; Bernardo and Smith, 1994, Section 5.4.1). This is, by the way, another very promising area for future research.

Question 44. Could you talk about general criticisms to non-subjective priors?

Answer. The major criticism usually comes from subjectivist Bayesians: the prior should be an honest expression of the analyst's prior knowledge, not a function of the model, specially if this involves integration over the sample space and hence violates the likelihood principle. I believe there are two complementary answers to this:

(i) *Foundational*: A non-subjective posterior is the answer to a *what if* question, namely what could be said about the quantity of interest given the data, if one's prior knowledge was dominated by the data; if the experiment is changed, or a different quantity of interest is considered, the non-subjective prior may be expected to change correspondingly. If subjective prior information is specified, the corresponding subjective posterior could be compared with the non-subjective posterior in order to assess the relative importance on the initial opinions in the final inference.

(ii) *Pragmatic*: In the complex multiparameter models which are now systematically used as a consequence of the availability of numerical MCMC methods, there is little hope for a detailed assessment of a huge personal multivariate prior; the naïve use of some 'tractable' prior may then hide important unwarranted assumptions which may easily dominate the analysis (see e.g., Casella, 1996, and references therein). Careful, responsible choice of a non-subjective prior is possibly the best available alternative.

Question 45. Could we end with some signpoints for those interested in pursuing this discussion at a more technical level?

Answer. The classic books by Jeffreys (1961), Lindley (1965) and Box and Tiao (1973) are a must for anyone interested in non-subjective Bayesian inference; other relevant books are Zellner (1971) and Geisser (1993).

The construction of non-subjective priors has a very interesting history, which dates back to Laplace (1812), and includes Jeffreys (1946, 1961), Perks (1947), Lindley (1961), Geisser and Cornfield (1963), Welch and Peers (1963), Hartigan (1964, 1965). Novick and Hall (1965), Jaynes (1968, 1971), Good (1969), Villegas (1971, 1977), Box and Tiao (1973, Section 1.3), Zellner (1977, 1986), Bernardo (1979), Rissanen (1983), Tibshirani (1989) and Berger and Bernardo (1989, 1992c) as some of the more influential contributions.

For a general overview of the subject, see Bernardo and Smith (1994, Section 5.6.2), Kass and Wasserman (1996), and references therein. Yang and Berger (1996) is a problem-specific (partial) catalog of the many non-subjective priors which have been proposed in the literature.

For some one specifically interested in reference priors, the original paper, Bernardo (1979) is easily read and it is followed by a very lively discussion; Berger and Bernardo (1989, 1992b) contain crucial extensions; Bernardo and Smith (1994, Section

5.4) provide a description of reference analysis at a textbook level; Bernardo and Ramón (1996) offer a modern elementary introduction.

References

Bayarri, M.J., 1981. Inferencia Bayesiana sobre el coeficiente de correlación de una población normal bivariante. Trab. Estadist. 32, 18–31.

Bayes, T., 1763. An essay towards solving a problem in the doctrine of chances. Phil. Trans. Roy. Soc London 53, 370–418 and 54, 296–325. Reprinted in Biometrika 45 (1958), 293–315.

Berger, J.O., 1985. Statistical Decision Theory and Bayesian Analysis. Springer, New York.

Berger, J.O., 1994. An overview to robust Bayesian analysis. Test 3, 5–124 (with discussion).

Berger, J.O., Bernardo, J.M., 1989. Estimating a product of means: Bayesian analysis with reference priors. J. Amer. Statist. Assoc. 84, 200–207.

Berger, J.O., Bernardo, J.M., 1992a. Ordered group reference priors with applications to a multinomial problem. Biometrika 79, 25–37.

Berger, J.O., Bernardo, J.M., 1992b. On the development of reference priors. In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M., (Eds). Bayesian Statistics 4, Oxford University press, Oxford. pp. 35–60 (with discussion).

Berger, J.O., Bernardo, J.M., 1992c. Reference priors in a variance components problem. In: Goel, P.K., Iyengar, N.S. (Eds.). Bayesian Analysis in Statistics and Econometrics. Springer, Berlin, pp. 323–340.

Bernardo, J.M., 1979. Reference posterior distributions for Bayesian inference. J. Roy. Statist. Soc. B 41, 113–147 (with discussion). Reprinted in: Polson, N.G., Tiao, G.C. (Eds.), 1995. Bayesian Inference. Edward Elgar, Brookfield, V.T, pp. 229–263.

Bernardo, J.M., 1981. Reference decisions. Symposia Math. 25, 85–94.

Bernardo, J.M., 1996. Bayesian hypothesis testing: a reference analysis. Workshop on Default Bayesian Statistical Methodology, Purdue University, 1–3 November, 1996.

Bernardo, J.M., Ramón, J.M., 1996. An elementary introduction to Bayesian reference analysis. Tech. Report, Universitat de València, Spain.

Bernardo, J.M., Smith, A.F.M., 1994. Bayesian Theory, Wiley, Chichester.

Box, G.E.P., Tiao, G.C., 1973. Bayesian Inference in Statistical Analysis. Addison-Wesley, Reading, MA.

Casella, G., 1996. Statistical inference and Monte Carlo algorithms. Test 5, 249–340 (with discussion).

Cifarelli, DM., Regazzini, E., 1987. Priors for exponential families which maximize the association between past and future observations. In: Viertl, R. (Ed.). Probability and Bayesian Statistics. Plenum, London, pp. 83–95.

Clarke, B., Barron, A.R., 1994. Jeffreys' prior is asymptotically least favourable under entropy risk. J. Statist. Planning Infer. 41, 37–60.

Consonni, G., Veronese, P., 1989. Some remarks on the use of improper priors for the analysis of exponential regression problems. Biometrika 76, 101–106.

Datta, G.S., Ghosh, J.K., 1995. On priors providing a frequentist validity for Bayesian inference. Biometrika 82, 37–45.

Datta, G.S., Ghosh, M., 1996. On the invariance of noninformative priors. Ann. Statist. 24, 141–159.

Dawid, A.P., 1983. Invariant prior distributions. In: Kotz, S., Johnson, N.L., Read, C.B. (Eds.). Encyclopedia of Statistical Sciences 4, Wiley, New York, pp. 228–236.

Dawid, A.P., Stone, M., Zidek, J.V., 1973. Marginalization paradoxes in Bayesian and structural inference. J. Roy. Statist. Soc. B 35, 189–233 (with discussion).

De Finetti, B., 1937. La prévision: ses lois logiques, ses sources subjectives. Ann. Inst. H. Poincaré 7, 1–68. Reprinted in 1980 as: Foresight; its logical laws, its subjective sources. In: Kyburg, H.E., Smokler, H.E., (Eds.). Studies in Subjective Probability. Dover, New York, pp. 93–158.

Diaconis, P., 1988. Recent progress on de Finetti's notion of exchangeability. In: Bayesian Statistics 3, Bernardo, J.M., DeGroot, M.H., Lindley, D.V., Smith, A.F.M. (Eds.). Oxford University Press, Oxford, pp. 111–125 (with discussion).

Efron, B., 1993. Bayes and likelihood calculations from confidence intervals. Biometrika 80, 3–26.

Efstathiou, M., 1996. Some Aspects of Approximation and Computation for Bayesian Inference. Ph.D. Thesis. Imperial College, London, UK.

Fishburn, P.C., 1981. Subjective expected utility: a review of normative theories. Theory Decision 13, 139–199.

Fishburn, P.C., 1986. The axioms of subjective probability. Statist. Sci. 1, 335–358 (with discussion).

Geisser, S., 1993. Predictive Inference: an Introduction. Chapman & Hall, London.

Geisser, S., Cornfield, J., 1963. Posterior distributions for multivariate normal parameters. J. Roy. Statist. Soc. B 25, 368–376.

Ghosh, J.K., Mukerjee, R., 1992. Non-informative priors. In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.). Bayesian Statistics 4, Oxford University Press, Oxford, pp. 195–210 (with discussion).

Good, I.J., 1969. What is the use of a distribution? In: Krishnaiah, P.R. (Ed.). Multivariate Analysis 2, Academic Press, New York, pp. 183–203.

Hartigan, J.A., 1964. Invariant prior distributions. Ann. Math. Statist. 35, 836–845.

Hartigan, J.A., 1965. The asymptotically unbiased prior distribution. Ann. Math. Statist. 36, 1137–1152.

Hartigan, J.A., 1966. Note on the confidence prior of Welch and Peers, J. Roy. Statist. Soc. B 28, 55–56.

Hartigan, J.A., 1983. Bayes Theory. Springer, Berlin.

Heath, D.L., Sudderth, W.D., 1978. On finitely additive priors, coherence and extended admissibility. Ann. Statist. 6, 333–345.

Heath, D.L., Sudderth, W.D., 1989. Coherent inference from improper priors and from finitely additive priors. Ann. Statist. 17, 907–919.

Hewitt, E., Savage, L.J., 1955. Symmetric measures on Cartesian products. Trans. Amer. Math. Soc. 80, 470–501.

Jaynes, E.T., 1968. Prior probabilities. IEEE Trans. Systems, Sci. Cybernet 4, 227–291.

Jaynes, E.T., 1971. The well posed problem. In: Godambe, V.P., Sprott, D.A., (Eds.). Foundations of Statistical Inference, Holt, Renehart and Winston, Toronto, pp. 342–356 (with discussion).

Jeffreys, H., 1946. An invariant form for the prior probability in estimation problems. Proc. Roy. Soc. A 186, 453–461.

Jeffreys, H., 1961. Theory of Probability, 3rd ed. Oxford University Press, Oxford.

Kashyap, R.I., 1971. Prior probability and uncertainty. IEEE Trans. Inform. Theory 14, 641–650.

Kass, R.E., 1989. The geometry of asymptotic inference. Statist. Sci. 4, 188–234.

Kass, R.E., 1990. Data-translated likelihood and Jeffreys' rule. Biometrika 77, 107–114.

Kass, R.E., Wasserman, L., 1996. The selection of prior distributions by formal rules. J. Amer. Statist. Assoc. 91, 1343–1370.

Laplace, P.S., 1912. Théorie Analytique des Probabilités. Paris, Courcier.

Lindley, D.V., 1956. On a measure of information provided by an experiment. Ann. Math. Statist. 27, 986–1005.

Lindley, D.V., 1961. The use of prior probability distributions in statistical inference and decision. In: Neyman, J., Scott, E.L. (Eds.). Proc. 4th Berkeley Symp. 1, University California Press, Berkeley, pp. 453–468.

Lindley, D.V., 1965. Introduction to Probability and Statistics from a Bayesian Viewpoint. vol. 2: Statistics. Cambridge University Press, Cambridge.

Lindley, D.V., 1996. Some comments on Bayes Factors. J. Statist. Planning and Inference, to appear.

Mukerjee, R., Dey, D.K., 1993. Frequentist validity of posterior quantiles in the presence of a nuisance parameter: Higher order asymptotics. Biometrika 80, 499–505.

Nicolau, A., 1993. Bayesian intervals with good frequentist behaviour in the presence of nuisance parameters. J. Roy. Statist. Soc. B 55, 377–390.

Novick, M.R., Hall, W.K., 1965. A Bayesian indifference procedure. J. Amer. Statist. Assoc. 60, 1104–1117.

Peers, H.W., 1965. On confidence points and Bayesian probability points in the case of several parameters. J. Roy. Statist. Soc. B 27, 9–16.

Perks, W., 1947. Some observations on inverse probability, including a new indifference rule. J. Inst. Actuaries 73, 285–334 (with discussion).

Renyi, A., 1962. Wahrscheinlichkeitsrechnung. Deutscher Verlag der Wissenschaften, Berlin. English translation in 1970 as: Probability Theory. Holden-Day, San Francisco, CA.

Rissanen, J., 1983. A universal prior for integers and estimation by minimum description length. Ann. Statist. 11, 416–431.

Savage, L.J., 1954. The Foundations of Statistics, 1972. 2nd ed. Wiley, New York. Dover, New York.

Shannon, C.E., 1948. A mathematical theory of communication. Bell System Tech. J. 27 379 423 and 623–656. Reprinted in: Shannon, C.E., Weaver, W., (Eds.). The Mathematical Theory of Communication. University of Illinois Press, Urbana, IL. 1949.

Smith, A.F.M., 1981. On random sequences with centred spherical symmetry. J. Roy. Statist. Soc. B 43, 208–209.

Stephens, D.A., Smith, A.F.M., 1992. Sampling-resampling techniques for the computation of posterior densities in normal means problems. Test 1, 1–18.

Stein, C., 1959. An example of wide discrepancy between fuducial and confidence intervals. Ann. Math Statist. 30, 877–880.

Stone, M., Dawid, A.P., 1972. Un-Bayesian implications of improper Bayesian inference in routine statistical problems. Biometrika 59, 369–375.

Tibshirani, R., 1989. Noninformative priors for one parameter of many. Biometrika 76, 604 608.

Villegas, C., 1971. On Haar priors. In: Godambe, V.P., Sprott, D.A., (Eds.). Foundations of Statistical Inference. Holt, Rinehart and Winston, Toronto, pp. 409 414 (with discussion).

Villegas, C., 1977. Inner statistical inference. J. Amer. Statist. Assoc. 72. 453–458.

Wasserman, L., 1991. An inferential interpretation of default priors. Tech. Report. 516. Carnegie Mellon University, USA.

Welch, B.L., Peers, H.W., 1963. On formulae for confidence points based on intervals of weighted likelihoods. J. Roy. Statist. Soc. B 25, 318–329.

Yang, R., 1995. Invariance of the reference prior under reparametrization. Test 4, 83 94.

Yang, R., Berger, J.O., 1996. A catalog of noninformative priors. Tech. Report. Purdue University. USA.

Ye, K., Berger, J.O., 1991. Non-informative priors for inferences in exponential regression models. Biometrika 78, 645 656.

Zellner, A., 1971. An Introduction to Bayesian Inference in Econometrics.. Wiley, New York. Reprinted in 1987, Krieger, Melbourne, FL.

Zellner, A., 1977. Maximal data information prior distributions. In: Aykac, A., Brumat, C. (Eds.). New Developments in the Applications of Bayesian Methods. North-Holland, Amsterdam, pp. 211 232.

Zellner, A., 1986. On assessing prior distributions and Bayesian regression analysis with g-prior distributions. In: Goel, P.K., Zellner, A. (Eds.). Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti, North-Holland, Amsterdam, pp. 233 243.

Comments on paper by J.M. Bernardo

D.R. Cox

Department of Statistics and Nuffield College, Oxford

It is very helpful to have Professor Bernardo's account of his important work on reference priors. The unusual format helps clarify the key issues involved.

I have one small historical comment, a major point of disagreement, and two questions.

The historical point is that the answer to question 17 might lead to an underestimation of the effect of Jeffreys's work. He was writing on these topics from before 1930 and had an important but relatively amicable published disagreement with R.A. Fisher in the 1930s. In the period of intense interest in foundational questions in the 1950s I know from first-hand experience that such influential North American workers as A. Birnbaum, D.A.S. Fraser, L.J. Savage, and J.W. Tukey were all familiar with Jeffreys's ideas; also the important work on Bayesian econometrics by Zellner builds on Jeffreys's ideas, I believe. It is of course true that Jeffreys's work on probability and inference was only a small proportion of his total scientific contributions.

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 $X_i \sim \mathrm{N}(\mu_i, 1)$ independently, the reference posterior distribution for $\phi := \sum \mu_i^2$ 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 "$\boldsymbol{\mu} \in A$", when $A = \{\boldsymbol{\mu}: \sum \mu_i^2 \leqslant k\}$? We get different answers depending on whether we use the reference prior for μ 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 $\sum \boldsymbol{\mu}_i^2$? 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 ζ of the parameter, with the density $p(z|\zeta)$ of z depending only on ζ, priors may be found for which the marginal posterior for ζ 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 ζ 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.

Additional references

Berger, J.O., Wolpert, R.L., 1984. The Likelihood Principle. IMS Lecture Notes – Monograph Series, vol. 6. Institute of Mathematical Statistics, Hayward, CA.

Dawid, A.P., Stone, M., Zidek, J.V., 1973. Marginalization paradoxes in Bayesian and structural inference. J. Roy. Statist. Soc. B 35, 189–233 (with discussion).

Dawid, A.P., Stone, M., Zidek, J.V., 1996. Critique of E.T. Jaynes's 'Paradoxes of Probability Theory'. Research Report 172, Department of Statistical Science, University College London. Available at http://www.ucl.ac.uk/Stats/research/abstracts.html.

Stone, M., 1982. Review and analysis of some inconsistencies related to improper priors and finite additivity. In: Løs, J., Pfeiffer, H. (Eds.), Proc. 6th Int. Congr. of Logic, Methodology and Philosophy of Science. North-Holland, Amsterdam, pp. 413–426.

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 coherent and free of the marginalization paradox; see Heath and Sudderth (1978) and Dawid et al. (1973). We have shown, Datta and Ghosh (1995), that in all the common problems the one parameter at a time reference prior is the right-invariant Haar measure and so has desirable properties but we do not known if this always the case. The Jeffreys' prior is the left-invariant Haar measure and so is known to give rise to incoherence and marginalization paradox. The reference prior based on parameters taken in groups suffers from the same problems.

We also show that if the maximal-invariant parameter is one-dimensional, then a suitable reference prior with the maximal invariant as the parameter of interest is always probability matching for this parameter. In all examples the one parameter at a time reference prior is this suitable reference prior.

In the absence of invariance under a group, a reference prior need not be probability matchings; see, for example, Ghosh and Mukherjee (1992) or Ghosh (1994).

In all examples the reference posteriors have been found to be proper. But no general result is known. Here is a simple counterexample. Suppose X_i's are i.i.d. and $X_i = 0$ with probability half and $\mathrm{N}(\theta, 1)$ with probability half. Here the Jeffreys' and reference prior are same and constant. The posterior is improper if $X_1, X_2, \ldots, X_n$ are all equal to zero. We have used such models in a geological mapping problem recently. I understand a very general sufficient condition for the posterior to be proper is now known for the Jeffreys' prior.

I have also two comments on the Professor Bernardo's (1979) innovative procedure. The first is about the functional and its asymptotic maximization. Asymptotically, the posterior is nearly degenerate. So the idea of comparing it with the prior is somewhat odd – to adapt a phrase of D. Basu in a different context, it is like comparing a mouse and an elephant (relative to the mouse). Inspite of this, the asymptotics gives sensible results because of certain mathematical properties of the functional used. Asymptotically, it breaks into a large constant term free of the prior and a sensible Kullback–Leibler term free of n – for details see Ghosh and Mukerjee (1992) or Ghosh (1994). If we ignore the large constant term, we are effectively comparing the prior of θ with the posterior for $\sqrt{n}(\theta - \hat{\theta})$ (in the regular case). This seems odd because, to carry on with our metaphor, we are inflating the mouse to make it comparable with an elephant.

Professor Bernardo had two very innovative ideas in his 1979 paper. The first was the introduction of the Lindley functional as a function of priors (rather than experiments as Lindley had intended). The second is the grouping of parameters and the construction of prior step by step. My second comment concerns the second idea. To fix ideas, suppose we have two one-dimensional parameters (θ_1, θ_2), arranged in order of importance. With usual notations about Fisher information, the Berger–Bernardo algorithm for the regular case essentially takes a geometric mean of $1/\sqrt{I^{11}}$ with respect to a measure with density $\sqrt{I_{22}}$. It is shown in Ghosh and Mukerjee (1992) and Ghosh (1994) that if, we take Bernardo's idea of taking parameters in groups but instead of maximizing his functional match posterior and frequentist probability, then one of the probability matching options is to choose a harmonic mean of $I/\sqrt{I^{11}}$ with respect to $\sqrt{I_{22}}$. This second idea of Professor Bernardo coupled with a new method of getting non-informative priors via limit of uniform distributions over finite approximating sieves in Ghosal, et al. (1996) leads to an arithmetic mean of $1\sqrt{I^{11}}$ with respect to $\sqrt{I_{22}}$. Computation by all three methods give rise to the same prior for the group models of Datta and Ghosh (1996) where Datta and I could carry out the calculations. I checked they are also identical for a multinomial with three classes. It would be interesting to know which of Professor Bernardo's two ideas is more important for the success of the reference posteriors.

To sum up, the basic criterion remains somewhat mysterious and, even though all – I repeat all – applications are so attractive, we still do not have general results on internal or external validation (except the invariance under reparameterization). More work remains to be done to settle these issues. In addition, there are challenging problems for studying reference priors and posteriors when the number of parameters is large or depends on n as in Neyman–Scott problems.

I finally turn to the entirely different issue of truncating an improper prior to a suitably large compact set. As Professor Bernardo says, one reason why this is not done is that analytic computations will then become impossible. However, if the posterior without truncation is improper should one truncate to get a proper posterior? The answer would seem to depend on a subjective but constructive assessment of how large the effective parameter space might be and stability of the inference with respect to perturbations of it. For example, for a real-valued parameter θ_i suppose we knew $|\theta| \leqslant 4$ and the posterior does not change much if we truncate the improper prior to $|\theta| \leqslant k$ where $k \sim 4$. Then truncation seems all right. However, if these conditions are not met, truncation would seem to be inappropriate.

Additional references

Ghosh, J.K., 1994. Higher Order Asymptotics. Institute of Mathematical Statistics and American Statistical Association.

Datta, G., Ghosh, J.K., 1995. Noninformative prior for maximal invariant parameter in group models. Test, 95–114.

Ghosal, S., Ghosh, J.K., Ramamoorthi, R.V., 1996. Noninformative priors for infinite dimensional problems via sieves and packing numbers, submitted.

Some comments on “Non-informative priors do not exist”

Dennis Lindley

Minehead, Somerset, England

Most statistical situations begin with a question, and data that bear on that question. Is the drug effective, and the results of a comparative trial. What is the extent of the effect, and an experiment to measure it. The usual approach is to model the set-up by adding to the data x a model which specifies a class of probability distributions for the data, ordinarily dependent on parameters θ and λ, $p(x \mid \theta, \lambda)$. Here θ is related to the question, for example, a measure of the effect, and λ is a nuisance parameter introduced to simplify the probability structure of the model. The Bayesian solution is to introduce a probability distribution that reflects one's knowledge of the parameters before the data were at hand, the so-called prior distribution, $p(\theta, \lambda)$, and then to calculate $p(\theta|x)$ using the rules for the probability calculus. $p(\theta \mid x)$ expresses our opinion of the parameter of interest in the light of the data, the posterior distribution. This paragraph has attempted to describe the Bayesian *standpoint*, formulating a prior, incorporating a likelihood and passing to a posterior.

In order to perform the calculation that produces the posterior, it is necessary to introduce techniques that are often quite elaborate. As a result, it has become common to develop classes of models for which the techniques have been worked out. Thus we have linear models, hierarchical models, and so on. The statistician, when faced with the original question and the data, will usually select the model from one of these classes. All this is part of Bayesian *techniques*. The distinction between standpoint and technique has been emphasized by de Finetti (1974). In particular, he has noted the separation that often occurs between the two ideas, with unfortunate consequences.

This separation leads to many statisticians concentrating on the technique and ignoring the standpoint. The likelihood part of the model and its analysis becomes the centre of activity because of the complexity of the ideas that are required to provide the solution. This concentration was first observed amongst statisticians of the classical school. It has unfortunately spread to the Bayesians who have found it hard to shrug off the indiscretions of the frequentists. Once Bayesians do this, they have trouble producing the prior. That should come from the question and the knowledge about it possessed before the data are at hand, i.e., from the standpoint. It is not part of the model, at least as understood by frequentists, and not part of the technique. What has happened is that, desperate to stay within the model, statisticians have tried to produce the prior from the model. Jeffreys was the first to make real progress here and reference priors, as so admirably expounded in this paper, are the latest, and most successful, attempt to achieve the aim of keeping it all within the likelihood part of the model.

My basic objection to default priors is that they are developed from the model and not from the knowledge possessed before the data are at hand. They ignore the practicalities of the situation, concentrating on technique, at the expense of the standpoint. For example, suppose that two statisticians are interested in the same question, expressed through knowledge of the same parameter θ, but that they have different data sets leading to different models $p_i(x_i|\theta, \lambda_i)$, $(i = 1, 2)$. Since the default prior is developed from the model distribution, the two statisticians will typically use different priors. And this, despite the fact that they initially had the same knowledge, only differing in the data available to extend it.

The effect of this is, to my mind, serious. The resolution of the statistics thereby forgets completely the reality of the problem. θ is treated, as de Fineitti said, as a Greek letter and its meaning ignored. Positive binomial trials have their default prior, no matter whether we are dealing with sex or the fall of coins. This prior differs from that of the negative binomial, thereby violating the likelihood principle which is central to the Bayesian approach, though illogically ignored by frequentists.

The resolution of the difficulty is clear; go back to the standpoint, recognize the reality of the Greek letter and think about things. The standpoint teaches us that data does not 'speak for itself'. It only speaks in relation to what we already know. Suppose John comes across a map that gives directions for getting to his house. It 'speaks' little to him, but to Jean, who wishes to visit John, it 'speaks' a lot. This has long been recognized with information, but not with data analysis.

There is much discussion in the literature about models. My own view, and I think that of de Finetti, is that the model is your description of the uncertainty present in you perception of the situation, the uncertainty being expressed in terms of probability. Thus, the prior is as much part of the model as is the likelihood. Once the model is settled in this complete form, technique is able to take over; but not until then. Technique cannot produce an opinion concerning the parameters, out of an opinion about the data.

Another objection to the default prior is the impropriety that often occurs. No one has ever held an opinion about something that is improper. A further objection is that an important feature of science is repeatability. Scientists reach their firm conclusions by repeating experiments. That is, they enter an experiment with the knowledge of the previous experiments. They put all the information together. Default priors deny this.

There is one defence of the reference approach which stands up. At least it produces a prior that is free from many objections; certainly from those possessed by the Jeffreys' prior. Within the Bayesian standpoint, it is necessary to determine a distribution $p(\theta, \lambda)$. For all but the simplest models, this is a formidable task especially since we do not have good procedures for assessing multivariate densities. Remember that the dimensionality of λ may be very high. Also it often does not have a simple, practical interpretation that is easy to think about. Reference priors may therefore be a sensible substitute until the necessary multivariate methods have been developed. But only until then. If I were a grant-giving body, I would prefer to give money for research into multivariate assessment, rather than into default concepts.

Reference

De Finetti, B., 1974. Bayesianism: its unifying role for both the foundations and the applications of statistics. Int. Statist. Rev. 42, 117–130.

Rejoinder

JOSÉ M. BERNARDO

Universitat de València, Spain

This is a fine set of discussions, and I am very grateful to their authors for their illuminating contributions to our understanding of a very polemic topic. Predictably, 'truth' is partially perceived by some discussants from a different perspective than mine, but their refreshingly sincere attitude is bound to help to clarify some of the more relevant issues. I will try to answer individually the queries which have been raised.

Reply to Professor Cox

I totally agree with Professor Cox on the importance of Jeffreys' work; not only did he pioneer a successful use of non-subjective prior distributions, but he produced a rule which, in the regular one-parameter case – the only case for which he strongly recommended its use – it is still regarded as 'the' appropriate solution. A scholarly account of his developments previous to his 1946 famous paper would certainly be very welcome by the statistical community.

It would have been a surprise if Professor Cox gave a foundational arguments the considerable weight Bayesians believe they deserve and, although the topic is clearly too deep to be adequately dealt with here, I welcome the opportunity he gives me to expand on it: (i) I fail to see the need for different types of uncertainty: indeed, I see as one of the strengths of the Bayesian approach that its notion of probability encompasses its semantic use, as well as those related to symmetries or to replications; (ii) it is certainly true the axiomatics of rational behaviour typically included a clearly idealized assumption of the comparability of probabilities, but then no physicist would claim the ability to measure with infinite precision, and yet physical measurements are assumed to be real numbers: I believe in the usefulness of a *normative* theory which assumes precise probabilities – taken with a large pinch of salt and a great deal of sensitivity analysis – (iii) the representation theorems are mathematical facts which only depend on the exchangeability assumption and do *not* depend on the particular view one might have on probability; although it is true that not all problems may be represented within this structure, *all statistical analysis which assume a random sample* from one model or another – and pragmatically those are an overwhelming majority – *are a particular case of the exchangeability structure* and, hence, they *require* a prior distribution for its logically correct analysis.

As Professor Cox remarks, I regarded as crucial the role played by the logarithmic concept of information in statistics; not only may it be used to define non-subjective priors, but much more generally, it provides a foundational basis to encompass statistical inference within decision theory (Bernardo, 1979), and a natural definition of the goodness of a probabilistic approximation in the very many instances where such a concept is required in statistics (Bernardo, 1987).

Finally, I would like to thank Professor Cox for drawing my attention to the problem posed by Box and Cox (1964) transformation families; this is indeed a very interesting problem and I am including the derivation of the corresponding reference posteriors within my 'to do' file.

Reply to Professor Dawid

Professor Dawid's main criticism to the use of reference priors is foundational: since the reference priors depend on the parameter of interest, focusing on different aspects of the problem would lead to different priors and, hence, to inconsistent results. The argument would indeed be devastating if reference priors were supposed to describe unique, personal, possibly 'diffuse' beliefs, but there are *not*! Reference analysis *must* be regarded as part of *sensitivity* analysis to the choice of the prior. Reference analysis clearly establishes that you cannot *simultaneously* have a prior which is minimally informative with respect to, say the μ_i's of a multinormal model *and* with respect to the sum of its squares, $\sum \mu_i^2$, hence the title of this paper. Consequently, a reference posterior *must* be regarded as the answer to a precise question on sensitivity: *if* I wanted to use a prior minimally informative with respect to ϕ, *then* $\pi(\phi|z)$ would encapsulate my inferences about ϕ. In a decision situation, where a unique prior must indeed be used, a class of reference priors for several parameters of interest may usefully be considered to help to understand the implications of the particular prior one is going to use, by precisely making explicit the possibly important judgements which such a prior implies about specific functions of the parameters.

Professor Dawid is certainly right when he mentions that it has not been *proven* that reference analysis always avoids the marginalization paradoxes; the problem is that the marginalization paradoxes are described by a set of examples, with no unifying theory on the general conditions which may produce them. The fact remains, however, that 25 years after they were discovered, no marginalization paradox has ever been encountered using reference priors, and that reference analysis is the only method to derive nonsubjective priors which successfully avoids the paradoxes.

Finally, I appreciated Professor Dawid's warning on the delicate aspects involved in the approximation of improper priors by a sequence of proper priors. This is precisely the reason behind the apparently involved definition of a reference *posterior* – as the limit, in the logarithmic divergence sense, of the sequence of posterior distributions obtained using Bayes theorem on a sequence of proper priors – rather than attempting a direct definition in terms of a limit of the priors themselves.

Reply to Professor Ghosh

As Professor Ghosh mentions, references priors in multiparameter settings may formally be defined with respect to any ordered set of parameter subgroups, but it is only by sequentially using the one-parameter solution – the 'one-at-a-time' reference prior – that satisfactory results are obtained. This could be expected both from the analysis of the two-parameter problems and from information-theoretical arguments, and has been precisely argued in Datta and Ghosh (1995) in problems which are invariant under an amenable group of transformations. This is why in multiparameter situations, I refer, to the one-at-a-time reference prior as 'the' reference prior (see, e.g., the answer to question 35).

The question of the conditions under which an improper prior leads to a proper posterior has not yet found a general answer. As Professor Ghosh mentions, in all examples reference posteriors, *given a minimum size sample*, have been found to be proper. It may reasonably be expected that general results could be established from the general definition of reference posteriors as limits of posteriors obtained from proper priors which I have just mentioned above, but I am not aware of any. In the example he mentions, the minimum size sample is clearly one non-zero observation, for zeros do not provide any information about θ; hence, as one would expect, the reference posterior of θ will not be proper until such a non-zero observation has been obtained, precisely indicating that there is nothing to be said about θ solely based on those data and the assumed model.

I very much welcome Professor Ghosh's insightful comments on the mathematics which, in the regular case, operate behind the maximization of the missing information required by the definition of a reference prior. However, I would like to stress that such a maximization may *also* be performed in non-regular cases – where Fisher's matrix may not even be defined – and that, even in regular cases, it is often simpler – as in the example above – to derive directly the form of the asymptotic posterior distribution from first principles, that it is to check the regularity conditions and obtain its form from Fisher's matrix.

The frequentist validation of reference posterior statements is important both theoretically – to guarantee that no inconsistencies may exist – and pragmatically – to establish bridges with non-Bayesian statisticians. As he mentions, more work remain to be done: I would specially like to draw attention to the need for further work with *small* samples; indeed available results only explain the good coverage properties of reference posterior regions in *asymptotic* conditions and, yet, simulations repeatedly suggest that, with continuous parameters, very good coverage properties are indeed obtained with *any* sample size.

Finally, the necessary approximation of open parameter spaces by convergent compact sequences in order to derive the reference distributions certainly requires further work. I believe one should always consider a probability model *endowed with an appropriate compact approximation* to its parameter space, which should then be kept *fixed*, via the appropriate transformations, for *all* inference problems considered within that model. A good candidate for such 'canonical' compact endowment could be the natural uniform approximation in the corresponding variance-stabilizing transformations (Bernardo, 1997).

Reply to Professor Lindley

Professor Lindley stresses the basic foundational arguments of Bayesian decision theory to argue that a prior is an expression of someone's beliefs and should therefore be independent of the model used. I certainly accept the formal argument, and I agree that in a decision-making situation, assessing a prior reflecting his or her knowledge is precisely what any decision-maker should *try* to do. However, a reliable direct probabilistic description of complex multivariate information is next to impossible, so that, when data may be expected to dominate the prior, one may be prepared to approximately describe one's prior information as minimally informative with respect to some specific aspects of the problem, if only as an insurance policy against a multivariate prior unsuspectedly overwhelming the information from the data in specific directions of interest. In that case, mathematics – the reference algorithm – could be used to transform this prior statement into a prior distribution (and the procedure could presumably be included within a project on multivariate assessment funded under Professor Lindley's guidance!). Moreover, as we move away from personal decision making and concentrate on scientific reporting, it is obvious to me that one is *forced* to perform some form of sensitivity analysis with respect to changes in the prior, with special emphasis in minimally informative prior situations, as I have tried to describe in my reply to Professor Dawid's comments; I believe that reference analysis does provide an appropriate mechanism for this type of work.

Professor Lindley further insists on the *physical*, real-world meaning of the parameters. I believe that the existence of such a meaning is the exception, not the rule; what is the physical meaning of, say, the many parameters with appear in a complex hierarchical log-linear model? The representation theorems guarantee that exchangeable observations may be regarded as a random sample from some probability model, whose parameters are *defined* as a limit of observables, and hence, unobservable themselves; direct assessment a probability distribution on an unobservable quantity cannot be made operational and, hence, a programme on subjective probability assessment about the parameters of a model is fraught with difficulties.

Professor Lindley also stresses the required propriety of the prior; as I have argued above, the whole theory of reference distributions is actually based on *proper* priors; it is only in the last step that limits are taken for mathematical tractability. It is a fact, however, that if the reference analysis is kept proper by working within the appropriate compact subsets and performing by simulation the required integrations, the results are numerically indistinguishable from those analytically available using the limiting form.

I can only agree with Professor Lindley's request for scientific repeatibility, but I disagree with his conclusions: the coverage probabilities of reference posterior regions have the kind of repeating properties that scientists often require, and this is something you cannot possibly obtain from subjective priors. With respect to sequential experimentation, it is less than obvious to me that one would always want to use as a prior the posterior from the last experiment; indeed, there is here a problem of temporal coherence: too many things would have typically changed between experiments for the same 'small world' contemplated by Bayes theorem to remain valid. Again, I would prefer an analysis of the present experiment from a 'what if', sensitivity analysis perspective, within which, exploring the consequences of assuming minimal information about the main quantity of interest may well provided the more useful results.

I would like to close by thanking again all the discussants for their thought provoking comments, and by thanking Dr. Irony and Professor Singpurwalla for providing so many questions and offering me this opportunity for a simulating debate.

References

Bernardo, J.M., 1987. Approximations in statistics from a decision-theoretical viewpoint. In: Viertel, R. (Ed.). Probability and Bayesian Statistics, Plenum, London, pp. 53–60.

Bernardo, J.M., 1979. Expected information as expected utility. Ann. Statist. 7, 686 690.

Bernardo, J.M., 1997. Comment to 'Exponential and Bayesian conjugate families: review and extensions', by E. Gutiérrez-Peña and A.F.M. Smith, Test 6, 70 71.

Box, G.E.P., Cox, D.R., 1964. An analysis of transformations. J. Roy. Statist. Soc. B 26, 211 252 (with discussion).