knowledge. I would prefer they devote more effort to the development of consistency (less pejorative) indices. For this task, their expertise is abundantly manifest. The exploration of whatever problems ensue for a quasi-Bayesian approach in the prediction of future observations is even more relevant. Too much importance has been vested in parameters and not enough attention paid to the prediction of observables. Here empirical methods can also serve. The investigator has the opportunity to either withhold a fraction of his observations or to secure new ones under the same conditions. Predictions about observables are then subject to public scrutiny in contrast to armchair speculation about hypothetical parameters that may exist only in the models of statisticians.

In conclusion, let me congratulate the authors on a genuinely elegant piece of work.

Professor D. A. S. Fraser (University of Toronto): The opportunity is most welcome to comment on the long, detailed and comprehensive paper by Dr Dawid, Professor Stone and Dr Zidek. Indeed the opportunities do not come easily to comment on the Bayesian viewpoint.

The paper presents two paradoxes, a large collection of examples and certain conclusions concerning statistical inference.

The “paradoxes” or anomalies are not new and have had substantial discussion elsewhere. The examples appear in puzzling profuseness suggesting the precise mechanisms to be illustrated have not been isolated. And the conclusions are unfounded or erroneous.

1. The paradoxes

The first “paradox” or anomaly is that a posterior integrated with respect to one parameter need not be a posterior from the component variable on which it is based. Or, in less Bayesian language, it is that a likelihood integrated with respect to one parameter need not be a likelihood or modulated likelihood from the component variable on which it is based (that is, a marginal likelihood or modulated marginal likelihood). From a non-Bayesian viewpoint there does not seem to be anything contradictory or paradoxical in this anomaly. For the common Bayesian practice of representing an unknown by a measure on a space certainly runs counter to the scientific method of representing an unknown by a space (the space of possible values); and of course there is the non-uniqueness of the measure: thus the foundations are such that one can hardly hope for reasonable consequences. What is surprising perhaps is that integration can work at all.

The essentials connected with the first anomaly have been discussed in two papers, Fraser (1971, 1972b).

The first of these papers notes that an integrated likelihood can give anomalous results and that in certain cases no choice of prior can give the appropriate marginal likelihood. Somewhat earlier Professor Lindley (1969) in his Biometrika review of The Structure of Inference had expressed confidence in this Bayesian integration procedure: “It is integration with respect to this prior that provides the marginal likelihood available for inferences (concerning the second parameter) . . .”. The paper was prepared as a rejoinder commenting on this and other inaccuracies in the Biometrika review, and was discussed and published as part of the Waterloo Symposium on Statistical Inference. It is of interest that the rejoinder was declined by Biometrika with reasons that remain unfathomable; and of parallel interest that the Bulletin of the American Mathematical Society solicited a rejoinder to a Bayesian review of the same book and published it with the review.

The second paper (Fraser, 1972b) discusses positive aspects connected with the anomaly —the determination of integration (and other) procedures that do give the appropriate marginal likelihood. It is of interest that this paper too had earlier contact with Biometrika and encountered a deeply committed Bayesian referee who reacted vehemently to second round changes made on the basis of his first round recommendations. Indeed it is not easy to comment on things bearing on the Bayesian viewpoint.
The second “paradox” or anomaly is how “two rights can make a wrong”; a better description is “two wrongs do not make a right”. In certain contexts the first anomaly can be avoided by using a right invariant prior. The second anomaly is that a right invariant prior on a group may be inconsistent with a right invariant prior on a larger (containing) group, with the consequent recurrence of the first kind of anomaly even with a right invariant prior. As noted above and in Fraser (1972a, 1973a) a right invariant prior may be a “wrong”; and it is not reasonable to expect two wrongs to be consistent or right.

The second anomaly has been discussed in a paper at the Third International Symposium on Multivariate Analysis in 1972 and published as Fraser (1973b).

This paper discusses the factorization of an invariant measure in terms of invariant measures for complementing subgroups, and it discusses the implications of this for Bayesian and structural inference. For the factorization let \( \theta \) in \( G \) be expressable as \( \beta \alpha \) where \( \beta \), \( \alpha \) are in complementing subgroups \( G_\alpha, G_\beta \), respectively; then

\[
d\nu(\theta) = d\nu(\beta \alpha) = d\nu(\alpha) \cdot \Delta^{-1}(\beta) \Delta_\alpha(\beta) \, d\nu(\beta)
\]

where \( \nu_\alpha \) and \( \Delta_\alpha \) are the right invariant measure and modular function for the group \( G_\alpha \).

Note One: An integration over \( \beta \) (right-coset integration) with respect to \( \nu \) is not in general an integration with respect to the right invariant measure \( \nu_\beta \). Note Two: An integration over \( \alpha \) (left-coset integration) with respect to \( \nu \) is essentially an integration with respect to the right invariant measure \( \nu_\alpha \) but it produces a resultant that has in general the additional factor \( \Delta^{-1}(\beta) \Delta_\alpha(\beta) \) multiplying the right invariant measure \( \nu_\beta \) for the remaining variable \( \alpha \).

These two notes describe the simple mechanisms that provide the underlying explanation for the “paradoxes” and the examples.

2. The examples

The authors’ paper presents a large number of examples in illustration of the two “paradoxes”, and the examples are supplemented by a rather detailed and extensive presentation of a group framework for the examples. All of this stands in rather sharp contrast to the short space needed to describe the “paradoxes”. A first reaction is that there are far too many examples. A more considered reaction is that the examples are an attempt at illustration and the precise mechanisms to be illustrated have not been isolated.

The first example considers several priors and shows that the right invariant prior avoids the “paradox”.

The next two examples are presented in terms of a right invariant for the full parameter and a “paradox” is then obtained for a component parameter; the “paradox” is avoided by suitably modulating the initial right invariant prior. The author’s third Example is not new and the anomaly is well known. For they could quote Sprott’s result (for example, as recorded in Fraser, 1964) that the posterior for the correlation coefficient from the progression group is Fisher’s correlation fiducial; and it has long been noted that that fiducial is not likelihood based.

The next two examples examine modulated right invariants for the full parameter and find that different modulations can be required for different component parameters.

The five examples present a mixed and rather unclear picture—the right invariant is a somewhat natural prior, but to examine a component parameter can require a modulation for the prior, and different modulations can be required for different components. The authors then propose that the confusion can be avoided by a restriction to proper priors. This is a strange proposal as a resolution of the difficulties—for it means in the interesting cases that one cannot eliminate a variable, and hence cannot go to the marginal likelihood. The difficulty vanishes because one chooses not to look at it!

There is in fact a simple group pattern underlying all five examples and the pattern is not revealed by the long and detailed group discussion in the authors’ Sections 2 and 3.
See Note One above in connection with Fraser (1973b): the right invariant measure factors as

\[ dv(\beta \alpha) = dv_1(\alpha) \cdot \Delta^{-1}(\beta) \Delta_2(\beta) \cdot dv_2(\beta), \]

and a right-coset integration with respect to \( \beta \) is not in general an integration with respect to the right invariant \( v_2 \) for the parameter being eliminated—because of the factor \( \Delta^{-1}(\beta) \Delta_2(\beta) \). All four component-parameter examples involve right-coset integration with respect to a natural group, and the introduction of a modulating factor is seen to be merely a device that eliminates the "extraneous" factor \( \Delta^{-1}(\beta) \Delta_2(\beta) \). Thus the substance underlying the first "paradox" and the five examples can be found in the right-coset factorization (Note One) of a right invariant measure.

In the framework of a structural model a right-coset integration on the parameter space corresponds to a left-coset integration on the error space; and it is known that such an integration violates the probability principles needed for the conditional distributions. Cautions in this regard were expressed in The Structure of Inference, and Dr Stone may recall a slighting remark on one of these cautions in his review of The Structure of Inference. The examples show a Bayesian need for an analogue of such cautions.

The next two examples do not have the simple group structure found with the first five examples. Accordingly there is no right invariant measure and it is not surprising that integrated likelihood cannot give the marginal likelihood for the component parameter. Each example can, however, be obtained from a larger model that does have group structure, and indeed obtained by a left-coset integration on the error space. The authors examine some aspects of this larger group model. The underlying mechanisms, however, are not revealed but again they can be found by the use of Note One. For this it is easier to translate from right cosets on the parameter space to the corresponding left cosets on the error space. Two levels of left-coset integration are involved: to obtain the distribution of the given variables; and to obtain the distribution for the marginal variable. No group step exists between the levels and accordingly there is no right invariant measure. Thus as would be expected the "paradox" occurs.

3. The structural example

The authors consider the bivariate normal as generated by a rotation and a positive lower triangular transformation from standard normal variables. And they conclude "that the theory of structural inference is powerful enough to develop its own paradoxes without the assistance of improper Bayesians".

The multivariate form of this bivariate model has been examined in detail as the central example in the paper "Inference and redundant parameters" presented at the Third International Symposium on Multivariate Analysis in Dayton, June 1972 and published as Fraser (1973b). Whatever power the "theory" may have—some comment on this later—and contrary to the authors' assertion, no paradoxes or contradictions are involved in the analysis of this bivariate or multivariate structural model. For some comments on difficulties with a weaker model, see Fraser (1973d).

4. The conclusions

The authors' "paradoxes" and examples have drawn attention to the difficulties and inconsistencies connected with the use of left invariant or flat priors, difficulties that derive as we have seen from factorization properties of the invariant measures. The difficulties provide a moderate range of indeterminacy in the possible posterior distributions. The authors conclude "that more statisticians will be guided by the philosophy of Lindley and Smith (1972) and turn their attentions to the characterization of prior knowledge, rather than prior ignorance".

The question whether subjective priors, personal biases of investigators, should be used in the statistical analysis of data and a statistical model will not be raised here; comments may be found in Fraser (1972a). Rather we consider the range in which an
investigator could reasonably present a subjective prior. All indications are that this range would be far wider than that connected with the indeterminacy of the posterior in the flat Bayesian analysis.

It seems then that the authors, on facing the indeterminacy with flat priors, are in fact making a virtue of indeterminacy by their suggested use of subjective proper priors. Certainly Professor Lindley's use of normal priors leads to nice multivariate calculations and we are entitled to use our own priors—good luck with the integrations—but going to greater indeterminacy can hardly be taken seriously as a conclusion from the indeterminacies discussed in the authors' paper.

Concerning the structural approach the authors suggest "that the adjacent layers of the theory are in basic conflict". As has been emphasized in Fraser (1972a) the structural approach is primarily concerned with finding those portions of statistical analysis that follow necessarily from the data and the statistical model alone. With data and the classical model one obtains the likelihood function and its model. With data and the structural model one obtains a conditional distribution that separates into components corresponding to parameters of interest. Classical theory can then provide the small and usually immediate step to tests of significance, point estimates and confidence intervals. The data and the structural model also label possible parameter values onto the space of the conditional distribution just described; perhaps the Bayesians' feeling of exclusive rights to posterior distributions draws their attention to this aspect of structural analysis—after all, all the Bayesian has is posteriors. Up to this point there is no "theory", only the determination of necessary results or logical consequences; the term "structural analysis" seems appropriate to refer to this. The authors almost touch on this aspect in their remark "...must be conditioned by no more than the logical deduction immediately available from the data...". The problem of determining things that are necessary within statistical analysis is an important one that concerns us all. This is particularly so if we consider the large number of principles and criteria and reduction methods that have been introduced to statistical analysis in the last 30–40 years in order to get results. A reassessment to determine necessary results seems important to me.

The use of the structural model is related to the objective identification of sources of variation or error; see Fraser (1972a). For example, is there an identifiable source for the error affecting the first variable \( x_1 \) in the authors' example? It seems easy to treat all events with their corresponding probabilities on an equal basis, but the results of Basu concerning ancillary statistics show that this is premature. Some recent comments on the consequent difficulties for the standard ancillary concept have been discussed in Fraser (1973c).

Concerning the authors' suggestion then, we are not involved with "adjacent layers" of a "theory" but rather with two different statistical models and whether one or the other is the appropriate model in some application.

5. Some details

The authors comment on "the expression of ignorance by means of invariance..." while Fraser's...theory of structural inference makes it almost axiomatic". As remarked above most of structural inference is concerned with the necessary analysis of the structural and classical models, not with theory. Ignorance is not expressed there by means of invariance as the authors suggest; it is expressed by the set, the set of possible values for the unknown.

"...while Fraser's theory is somewhat controversial at the initial axiom level". Structural analysis is involved with a model and the necessary analysis of it. No axioms are involved beyond that needed for the mathematics used to describe the model; for example, to describe a probability space.

It seems unusual to include the proof without references for the standard invariance result: the distribution of the maximal invariant variable depends only on the maximal invariant parameter.