DISCUSSION

I.R. DUNSMORE *(University of Sheffield)*:

We heard in the discussions yesterday a prediction that Bayesian statistics would be dead by the end of the twentieth century. Many of us may still be around then to ensure that this will not be the case; and some of us are going about this task by concentrating attention on predictive distributions of future observations rather than on posterior distributions of parameters. A welcome step in the right direction therefore is this interesting and clearly presented paper by Professor DeGroot in which he attempts to model how a statistician proposes predictive distributions in a sequence of similar decision problems. Can we apply the method practically?

After hearing and assessing all the information at my disposal my initial (predictive) statement is that the theory is beautifully modelled but from the practical viewpoint I am dubious of its worth. Yet Professor DeGroot is a leading authority, and so following his doctrine of probabilism or probabiliorism that a doctrine recognized by a leading authority to be correct can be taken to be correct, the reality must be that I am subject to appreciable specification bias and that what has been presented lies well out in the tails of my initial predictive distribution. So I must learn to do better and reexamine the information available.

My main stumbling point is the separation of "previous information" from "tendency to mis-specify". Consider the problem S_1 in the location-shift model. After duly considering *all* his available information the statistician specifies a predictive distribution for X_1, say $N(M_1.T_1)$. Professor DeGroot argues that we could arrive at this distribution by modelling the statistician's behaviour as follows. He is just about to specify that X_1 is $N(m_1,r_1)$ when he remembers his specification bias θ_1, and so takes X_1 as $N(m_1+\theta_1/r_1^{1/2},r_1)$. This of course presumes that he can separate his information for assessing $N(m_1,r_1)$ from his specification bias information. With a prior $N(\mu_1,\tau_1)$ on θ_1 this distribution averages out to a predictive assessment of $N(M_1 = m_1+\mu_1/r_1^{1/2},T_1 = (\tau_1/(1+\tau_1))r_1)$. Now move on to problem S_2. The statistician will learn something from observing the outcome of S_1 about the type of specification errors he is likely to make and, consciously or subconsciously, he will use this knowledge about his own personality when he initially specifies the predictive distribution of X_2 as $N(M_2,T_2)$. Professor DeGroot then suggests the same procedure for modelling the learning by arguing that the statistician is just about to specify a $N(m_2,r_2)$ distribution for X_2 when he remembers his second specification bias θ_2, with prior $N(\mu_2,\tau_2)$, so that he gives his predictive distribution as $N(M_2 = m_2+\mu_2/r_2^{1/2},T_2 = (\tau_2/(\tau_2+1))r_2)$. But here θ_2 must measure in the learning process that statistician's "failure to correct sufficiently" in specifying m_2 and r_2. These seem to me to be much more nebulous quantities to deal with, and it is most doubtful if the assessment of m_2 and r_2 is now independent of the information for assessing θ_2.

For the location shift parameter Professor DeGroot models his procedure most elegantly, but I am much less convinced with the modelling of the scale or precision example of §6. There does not seem to me to be any intuitive appeal in the learning process $\alpha_2 = \alpha_1' + \delta, \beta_2 = \beta_1' + \delta$ when R_2 somehow measures the statistician's "failure to correct sufficiently". The problem presumably intensifies for the location and precision shift problem.

One final comment concerns the element of distribution dependence in the arguments presented. No observation will allow the statistician to move away from a normal distribution (or a Student distribution) for example. Might it not be that if 80% of your observations lie more than 5 standard deviations below the mean of your predicted distribution then skewed predictive distributions may be more appropriate.

Turning now to the second paper, my first task is to thank Professor Press for his interesting talk today. The Bayesian modelling of his qualitative controlled feedback ideas was clearly the next progression in his study of group judgment formulation - althought many here might presumably have expected it to be the first step. Professor Press has presented us with several different Bayesian approximate models via some elegant and intricate matrix manipulation.

My second task is not as easy. The question I wish to pose is "What does it all mean from a practical point of view?". The practical relevance and interpretation of the posterior distribution for $\mathbf{a}$ somewhat defeats me. For example, just consider the dimensionality of $\mathbf{a}$. This is an $hx1$ vector in the model (7) where $h = \Sigma_{\ell=1}^{\pi-1} h_\ell = \Sigma_{\ell=1}^{\pi-1} r \Sigma_{j=1}^{q} R_\ell(j)$. So if, for example, there are $r=5$ explanatory variables, $q=2$ questions, and $R_1(1)=4$, $R_1(2)=5$, $R_2(1)=6$, $R_2(2)=9$, then by the third stage h is already 120. Another point is that some of the reasons within the list $R_2(1)$, for example, will be contained in the list $R_1(1)$. Is it then necessary to have different $a_\alpha(1)$'s for stages 1 and 2 or could we allow the $p_n^{(\alpha)}(j)$'s in the model to account for the variability over n? (Although the notation is not explicit on this fact it seems that the a's are considered to vary with n).

It is also clear that the adequacy or otherwise of this ingenious model must be thoroughly investigated. One question on this count that I would like to pose is to ask if the model can cope with someone who "about turns" in his answer or opinion even with the same reasons, which after all is a common tactic of some committee members!

Turning briefly to the approximations to the posterior distributions I have one comment on the large-sample approximation of §3.3. Is it sensible to ignore the variable with the "largest" variance matrix? To an outsider it must look as though your conclusions will be more accurate than they should have been.

My third and final task is simply a plea. Please may we see this elegant statistical theory transformed into useful statistical practice.

S. GEISSER (*University of Minnesota*):

Professor DeGroot, no doubt manifesting his great flair for the sensibilities of our Spanish hosts, presented this paper because it implements what that renowned Spanish and American philosopher, Santayana, rather broadly intimated that those who ignore the data are doomed to repeat the mistakes of the past.

DeGroot has elegantly formulated a Bayesian apparatus that might serve to dampen and eventually avoid subjective biases. In some respects, he is more vitally concerned with the non-Bayesian aspects of the problem. He lists several possible alternative conclusions to be drawn from the fact that a subjectively predicted value is far from where expected:

1) The subjective predictive distribution is misleading (model doesn't fit).
2) A rare event has occurred because of an unusual observation (all we have to do is stick the conjunction "or" between (1) and (2) and we have Fisher's so-called logical disjunction).

What is left then is the last alternative which states:

3) There is no relation between the predictive distribution and the observed value.

I would call (3) the archsubjectivistic view, but I refuse to pay it much heed because if anything could persuade me to turn in my Bayesian credentials, it is this extreme view. And, of course, DeGroot is also too sensible to accept this. (What then is the point of a predictive distribution if it is to bear no relation to an observation?) At any rate, he claims to model what he terms the "behaviour" of the statistician. (Wasn't it Neyman who coined the term inductive behaviour?) It would appear that the accretions of the past are not so easily disposed of on the ash heap of history, and perhaps rightly so. In modelling the situation he finds he must explore the mind-boggling hyperworld of hyperparameters. But DeGroot is a slippery Bayesian and he refuses to assume an extreme position by not trying to n-guess himself, thus extricating himself from the hyperparametric thicket he has created.

He sensibly assumes that his specification error should be smaller the second time by a fraction of the first expected specification error -and so it goes, recursively that is.

There is also the hyperworld of γ itself, the DeGroot rate of learning parameter which is also subjected to a distribution and as decisions made and observations obtained he learns about his learning rate —and he learns about how he learns about his learning rate— a veritable underworld of learning similar to Fisher's underworld of probability.

It seems that this method of "learning" could easily be called the "DeGroot Opinion Processor and Evaluator" whose descriptive acronym aptly describes the learner. It is a smooth "rational" method that has none of the qualities of human learning at its best - inspiration, acuity, perception, and concentration - and is better suited for plodding, dull, and unimaginative automatons.

In the light of all this perhaps mildly unfair criticism, let me also give Professor DeGroot something less amorphous to which he can respond. Consider the predictive subjective model which is normal and unbiased, but may be more widely dispersed than it should be, whatever that may mean, then how important is it for the statistician who guesses only the mean value of his subjective distribution, and if so, how concerned should he be about having too large a subjective misspecification variance if he only will guess a single value?

Also, couldn't the statisticians' misspecification bias really be due to a misreading of auxiliary conditions which may affect the payoff of investment decisions, and instead of smoothly adjusting his parameters, he may want to radically reconsider his whole set-up after a few "bad" decisions.

In conclusion, let me say that I thoroughly enjoyed this paper as it compelled me to consider how important it is for a Bayesian to become involved in the human learning process and what a giant step DeGroot has taken in grappling with this problem and developing a point of view which is certainly not entirely orthogonal to the truth, if there is any here.

Professor Press is to be congratulated on his usual virtuoso performance in manipulating distributions of random matrices. But the refrain "What's it all about Alfie?" keeps coming to mind. If our good friend the new socialist mayor of Valencia wanted to resolve a pressing public policy issue -say the building of a Bayesian conference Center- why should he use Qualitative Controlled Feedback rather than having an open discussion and a popular referendum?

Another difficulty that I have previously pointed out about feedback procedures is the potential for misuse by a devious intermediary who would feedback false or slanted information in order to manipulate the outcome.

J.M. BERNARDO (*Universidad de Valencia*):

The nice mathematical properties associated to the combination of a normal model with the inverted Wishart distribution used by Professor Press in his prior specification have been exploited in a number of Bayesian papers. However, as he points out, this may be too restrictive. I would like Professor Press to expand on this point, making explicit the type of situations for which he feels this prior might be sensible, and commenting whether he knows of any real life applications. Information about possible interactive computer routines for this type of prior specification would be valuable.

A.P. DAWID (*The City University*)

Professor DeGroot is surely right to argue that the Bayesian should be ready to confront his internal probabilistic view of the world with some external reality, and to modify his view, rather than the real world, if there appears to be a conflict. One of the weaknesses of subjectivist theory, confined as it is in its comfortable coherent cocoon, is that it does not seem to make any formal allowance for such a confrontation. Something can be said, however.

Suppose that a weather forecaster has to make, each day, a statement of his probability of precipitation within a specified 12 hour period of the next 24 hours (Murphy and Winkler, 1977). He need not have any model in mind, but is merely stating his conditional probability of "rain tomorrow", given his whole knowledge today. Let us now consider all those days for which his forecast probability lay in the range, say, $1/3 \pm \epsilon$, and suppose that the number of such days is (conceptually) infinite. Then, using martingale theory, one can show that the limiting relative frequency of rain on such days lies, with probability one, in the same range $1/3 \pm \epsilon$. The probability referred to here is, of course, that corresponding to the forecaster's subjective opinions.

Similarly, if each day he gives a *credible interval* which he assesses to have, say, 50% probability of containing tomorrow's maximum temperature, then he should believe, with probability one, that in the limit 50% of such intervals will contain the true value.

Note particularly that the above theory does not require any assumption of independent or "unrelated" problems, merely that each forecast be made in the light of full knowledge of the outcomes of previous forecasts. So, in a sense, the Bayesian is out-frequencing the frequentist.

Now suppose that, in a very long sequence of such forecasts, only 30% of the forecaster's 50% credible intervals are covering their true values; then an event has occurred to which the forecaster assigned very low probability. It seems to me clear that the world is telling the forecaster that his Bayesian beliefs, coherent though they may be, are out of touch with reality. However this logic is squarely in the spirit of significance testing (or of Professor Box's contribution to this conference) and I cannot see how to justify it from the position of the self-contained subjectivist.

The above considerations apply for the forecaster's "true" probabilities. It is easy for him to cheat, by quoting probabilities in which he does not really believe, so as to appear "well-calibrated" (DeGroot, 1979). Moreover, even if his true probabilities are well-calibrated, this does not necessarily mean that they are "accurate" in all respects; and even if they are accurate, they may not be of much *substantive* value if the forecaster is a poor meteorologist.

Professor DeGroot is working in the following framework. The forecaster sets up a mathematical model which, he hopes, is an adequate approximation to his true internal beliefs, which he in turn hopes correspond, somehow, to the real world. But the real world says "Not so". So the forecaster replaces his initial model with a more complex one, which he hopes will lead to more "accurate" forecasts. Clearly the process can be iterated, and now bears a very close resemblance to Box's cycle of estimation and criticism.

But I believe there is a danger of falling into an infinite regress. However much we refine our subjective models, or learn about our learning process, the real world may still surprise us by throwing up events which we believe shouldn't occur. So in what sense, if any, have we improved our probability modelling?

W.H. DUMOUCHEL (*Massachussets Institute of Technology*):

Professor DeGroot should be thanked for tackling the somewhat taboo question: "What can a Bayesian do who is consistently wrong?" I would like to suggest another possible approach and solution.

Suppose the statistician encounters a sequence of trials in which it is necessary to predict a continuous variable S, after which the observation $X = x_n$ is made at the n^{th} trial. Let F_n be the statistician's predictive distribution function for X just before the n^{th} trial, and let

$$U_n = F_n(x_n) \quad ,$$

which is observable after the n^{th} trial. Now, if the statistician computes F_n correctly just before the n^{th} trial, the sequence, U_1, U_2, ... should be indistinguishable from a sequence of independent uniform variables on (0,1). If the predictive distribution F_n is being consistently computed incorrectly, then it may be that $\{U_n\}$ behaves like a sequence of i.i.d. variables from some unknown distribution G. A possible assumption is that G is a beta distribution with parameters α and β. Then the question of whether the statistician is consistently wrong in computing F_n for the predictive distribution of X_n boils down to an hypothesis about α and β, where the variable $U_n = F_n(X_n)$ has a Beta (α,β) distribution, i.e,

$$H\colon \alpha = \beta = 1$$
$$H\colon (\alpha,\beta) \neq (1,1) \quad .$$

This problem can be treated as a sharp null hypothesis problem. The statistician formulates a prior distribution on (α,β) with

Ω_0 = prior odds in favor of H
Ω_n = odds in favor of H conditional on $U_1, U_2,\ldots, U_n$.

Then the predictive distribution of U_{n+1} is $\hat{G}_n(u) = \int G(u;\alpha,\beta)dP_n(\alpha,\beta)$, where $G(u;\alpha,\beta)$ is the beta distribution function, and P_n is the posterior distribution of (α,β).

My proposed decision rule is then:

1. Make no corrections to inferences about X as long as $\Omega_n \geq 1$(or $\Omega \geq k$).
2. If $\Omega_n < 1$, then correct F_{n+1} to make U_{n+1} uniform. That is, $U_{n+1} = F^{*}_{n+1}(x) = \hat{G}_n(F_{n+1}(X))$. The 100 u percentile of x_{n+1} is $F^{-1}_{n+1}\ \hat{G}^{-1}_n(u)$. In general, for $k \geq 1$, replace F_{n+k} by $\hat{G}_n F_{n+k}$ in all inferences about X.
3. Whenever $\Omega_n < 1$ and step 2 above has been taken, start over with a new reassessment of $\Omega_0, P_0(\alpha,\beta)$, etc. corresponding to the new definitions of F and U.

If the assumption that $\{U_n\}$ is approximately a sample from some beta distribution is correct, then when n is large, $\hat{G}_n\, F_{n+1}$ will produce just the right correction to F_{n+1}, as simple calculations show. Of course it is no simple calculation to compute Ω_n or $\hat{G}_n$, which depend on the choice of prior $P_0(\alpha,\beta)$ for α and β when n is small, but less so when n is large.

Although this method poses computational problems, it is very general, being applicable to any continuous predictive situation, and it provides a method for simultaneously correcting for error in scale and location, since the two parameter (α,β) are available for the estimation of G, and even more general families could be used instead of the beta family. I hope to develop this method in future work.

S. FRENCH (*University of Manchester*):

I should like to comment on Professor Press' paper. It seems particularly important to emphasise a point that was clear from Professor Press' presentation at the conference, but not clear from his written paper. At least, I for one was misled. The methods of this paper are directed at the problem of gathering and summarising group opinion for a decision maker *exterior* to the group. They should not be considered as methods to help a group of decision makers reach consensus amongst themselves. That no such methods can exist should now be well known, Arrow (1963), Luce and Raiffa (1958), Patternaik (1978), French (1980). I say "should be" since I am aware that some decision analysts see their task of advising a group as one of generating a group probability distribution and a group utility function and then of advising the action with maximum group expected utility. Such analyses are unlikely to be rational in the Bayesian sense. Professor Press' methods appear tailor-made for such "irrational" analyses. I was glad to hear from his presentation that such a close fit was unintentional.

Turning now to the correct use of Professor Press' analysis, I am far from convinced that anonymity will lead to "objetivity". I personally judge a person's opinions and his reasons for holding those opinions against the background of his character. Moreover, I am aware that some of the best opinions are held without the holder being able to express why he holds them. Consider a firm taking advice from a group of experts within a research and development department. How will Professor Press' method assimilate the opinions of a man with a hunch. By definition he cannot articulate his reasons for his opinion. So his view will not communicate itself to the rest of the group. Yet, the rest may all agree "Old Charlie has a gut feeling for winning projects. If he says it's a winner then that's good enough for me". This example is contrived maybe, but I hope it makes my point. One gathers information through a group of experts rather than the literature, when it is clear that there are too many uncertainties involved for them to be objectively analysed. Thus one intentionally asks the panel to use their intuitive expertise. Yet this method concentrates their attention on that part of their judgemental process which they can articulate, but not necessarily directly upon the part for which they were employed.

It may well be that Professor Press does not see this method as being used to sample expert opinion, but rather a large population of consumers. His paper does indeed concentrate on an example where a city planning bureau surveys public opinion. (However, see Harman and Press (1978)). Here too, I am worried about the applicability of his methods. A sample survey is meant to be representative of the population sampled. Yet it is a basic property of qualitative controlled feedback that it changes the initial opinions of the group. So the output of a sample survey conducted by Professor Press' methods is unlikely to be representative of the opinions sampled. At the end of the analysis those in the sample will have thought about their position more carefully than the rest of the population. However desirable it is that public policy should be based upon well informed and well thought opinion, I suspect that politicians would rather base it upon opinion as it is.

I.J. GOOD (*Virginia Polytechnic and State University*):

Dr. DeGroot referred to the situation where there are two or more statisticians who have specified predictive distributions for X. I think that theory is directly related to the problem of how a single statistician can improve his judgement, namely by comparing a number of procedures for specifying priors as if they were provided by several statisticians. In other words he can, so to speak, split his personality. One of the ways of seeing which statistician is better at predicting is by means of the logarithmic payoff function which I advocated in 1951 (Good, 1952). If the probability or probability density of the observed value of X (say x) is $p(x)$, the logarithmic payoff function is of the form $a+b \log p(x)$. This is one of the payoff functions that encourages the statistician to be honest, and when comparing two statisticians the gain of the first over the second is proportional to $\log[p_1(x)/p_2(x)]$ (in a self-explanatory notation). This has a further justification; we can imagine that there is a demiurge with perfect judgement whose probability (density) is $p_0(x)$. When comparing a statistician with the demiurge we could imagine that we were trying to find out which of the two was the demiurge. Then $\log [p(x)/(p_0(x)]$ would be the weight of evidence in favour of the statistician's *being* the demiurge. We could imagine that we score *each* statistician in this way against the demiurge. Then the gain of statistician 1 over statistician 2 would be

$$\log [p_1(x)/p_0(x)] - \log [p_2(x)/p_0(x)] = \log [p_1(x)/p_2(x)]$$

so we don't need to know $p_0(x)$ for trying to decide which of the two is better. If there is a true probability density, then the expected advantage of 1 over 2 is

$$\int p_0(x) \log [p_1(x)/p_2(x)]dx$$

I have a comment concerning Dr. Box's comment. Dr. Box said that the observed ordinate $p(x)$ of the probability density should not be compared with the density at the mode, and so he asked about using the tail-area probability. If instead you compare $p(x)$ with the *average* value of the probability density then you would be using Warren Weaver's surprise index. I generalized Weaver's surprise index to a continuum of indexes in Good (1953, 1956) where I invented what has been called Rényi's generalized entropy. (Perhaps it should be attributed to Good). A special case of the generalization is $\int p(y) \log p(y)dy - \log p(x)$.

D.V. LINDLEY (*University College London*):

An alternative way of handling this problem is to suppose that the statistician is observed by a totally coherent person who takes the statistician's views and updates them in the light of experience with similar outcomes. This has been explored by Lindley, Tversky and Brown, (1979) . Equivalently, the statistician can think of his incoherent, natural self being monitored by a coherent person inside him. It is not obvious to me which approach is preferable but ours does appear to avoid the need for assumptions like (4.1), This conference has been dominated by technical papers and it is a real pleasure to welcome this thoughtful paper which tackles an important problem.

A. ZELLNER (*University of Chicago*):

In connection with DeGroot's suggested adaptive learning approach, consider two hypotheses regarding a parameter θ, namely $H_1{:}\theta = \theta_0$, a given value and $H_2{:}\theta \neq \theta_0$. If we have posterior probabilities for these hypotheses, p_1 and $1\text{-}p_1$, the optimal (relative to a symmetric loss function) estimate of θ is $\hat{\theta} = p_1\theta_0 + (1\text{-}p_1)\bar{\theta}$, where $\bar{\theta}$ is the posterior mean of θ under H_2. $\hat{\theta}$ can be equivalently expressed as $\hat{\theta} = \theta_0 + (1\text{-}p_1)(\bar{\theta} - \theta_0)$ and it is seen that $1\text{-}p_1$ is an "adjustment coefficient" that is data dependent. Similarly, when we consider two alternative models with posterior probabilities, p_1 and $1\text{-}p_1$, the optimal point prediction is $\hat{y}=p_1\hat{y}_1+(1\text{-}p_1)\hat{y}_2=\hat{y}_1+(1\text{-}p_1)(\hat{y}_2\text{-}\hat{y}_1)$ where $\hat{y}_1$ and $\hat{y}_2$ are means of the predictive distributions for the two models. Again $1\text{-}p_1$ appears as a data dependent adjustment coefficient. These traditional Bayesian procedures incorporate adaptive learning and thus there may be no need for an alternative learning model such as proposed by DeGroot.

REPLY TO THE DISCUSSION

M.H. DEGROOT (*Carnegie-Mellon University*):

I am grateful to all the discussants for their comments and their appreciation of the general problem that I am trying to attack in this paper. Both Dr. Dunsmore and Prof. Geisser comment on possible shortcomings and difficulties with the models that I have presented. As Dr. Dunsmore suggests, I should extend my models to cover shape misspecification and to include skewed distributions.

Prof. Geisser says that the learning process in my model doesn't provide for inspiration and is "better suited for plodding, dull, and unimaginative automatons". At first I thought that he was criticizing my model, but then I realized that he was actually pointing out that my model appropriately describes the learning process of most statisticians. More seriously, learning proceeds in my models neither too slowly nor too quickly, but at just the right rate, i.e., Bayesianly. If one wishes to allow for the "inspiration" of changing models based on the data, then these possible changes must be, and can be, incorporated into a supermodel.

I agree with these discussants—we do need better models. But I believe that the development of such models should go hand-in-hand with the necessary psychological modeling.

In answer to a question raised by Professor Geisser, precision misspecification is relevant, even if the statistician is only going to use the mean of his predictive distribution as his predicted value. Although the statistician's predicted values may be unbiased, he will find that they tend to be much closer to, or much farther from, the correct values than he anticipated. Incidentally, it would be nice if one fringe benefit of this work was to introduce colorful terms like "bias" and "unbiasedness" into Bayesian statistics and reclaim them from sampling theory statistics where they have been wasted on useless concepts.

As Professor Lindley suggests, I am sure that there are times when it can be helpful to suppose that an incoherent statistician has a shadowy coherent alter ego looking over his shoulder or a tiny coherent elf somewhere inside him struggling to emerge. But two aspects of my work should be emphasized: First, the statistician may be biased, but he is coherent. Second, an important purpose of the models is to reduce and ultimately eliminate the need for the statistician to carry on any dialogue with himself.

Professor Good also suggests that the statistician can split his personality and see which personality makes the best predictions. He should then, I suppose, adopt that personality (at least whenever he must make a prediction). Professor Good suggests the use of scoring rules to see which personality is doing best. One difficulty with the use of any particular scoring rule is that it must be assumed that the statistician's expected utility function is simply his expected total score over a sequence of predictions. But if the different personalities have different subjective probabilities, wouldn't they also have different utility functions? Again, I emphasize that one purpose of my models is to eliminate split personality, which is a step toward better mental health as well as better statistics.

Professor Zellner is correct in suggesting that the standard Bayesian methodology for choosing among different models may be adequate in describing the learning process in many situations. The essence of my models, however, is to carry over into future problems what we have learned in earlier problems about *how* to specify prior distributions. That idea seems to me to be new.

Professor Dawid makes several interesting and valid points. I do believe, however, that when the world tells the forecaster that his beliefs are out of touch with reality, the forecaster can recognize this message and make adjustments wholly within the Bayesian framework. He does not need to use the logic or methodology of significance testing, although a forecaster whose faith is weak would be tempted to do so. It is true that in order to make these adjustments, the forecaster must go to a hierarchical model with perhaps a large number, possibly even an infinite number, of levels. But if, as Professor Good states in his paper at this conference, the hyperparameters at the higher levels matter less and less, then he will have improved his forecasts.

Professor DuMouchel proposes a clever new model, and avoids the methodology of significance testing by carrying out a Bayesian test of his hypotheses at each stage. The model promises to be fairly comprehensive and clearly warrants further study, development, and application.

S.J. PRESS (*University of California, Riverside*):

The qualitative controlled feedback (QCF) data collection protocol is a procedure for collecting information of various kinds from a group; the information can be used and analyzed in a variety of ways. This broad base of applicability is one of the greatest assets of the approach. The procedure can be used for example, merely to collect arguments and justifications in favor of one policy or another that has been advocated. Group members can bring to bear arguments based upon information each of them has separately, and information they have generated together as a group, and they can also argue various positions on the basis of information they might not have originally, but later are exposed to, and they can evaluate it in a meaningful way. Group members may differ in the amounts of information they have available, the type of information they have available, and in their ability to verbalize arguments using this information. They will differ in their experience level, intellect, intuitive ability, and expertise. They will share however, a large base of intellect, rationality, and information. The variation in opinion, after several rounds of QCF, is in itself a measure of the uncertainty or lack of knowledge surrounding the situation. The results are therefore very meaningful even when consensus is not achieved. Many applications will involve no more than just a collection of arguments arrived at after several iterations of the QCF process. Such arguments may be useful for assessing risks and for evaluating a complicated situation. In other applications it may be useful to develop quantitative information about some important questions using opinion and arguments generated by the group. In these cases, the absolute answers may be of fundamental importance, or what may really be of interest is the change, over time, in the group's perception of the basic answers to the fundamental questions. In these kinds of applications it is useful to use the QCF procedure with a quantitative base. Finally, in still other applications, it may be useful to use a model, such as the one developed in the paper, for predicting the next round's quantitative outcome based upon earlier developed information. With these prefatory remarks I now turn to the thoughtful questions raised by participants at the Bayesian Conference.

Professor Bernardo raised the question of how one actually uses an inverted Wishart distribution in practice. An then, how does one use the more complicated $_rF_q$ generalized distribution discussed in the paper? This question is an important one from the point of view of practical applications of Bayesian methods in general, because the inverted Wishart distribution family is the one most often proposed as the family of natural conjugate priors that should be used for scale parameters. The inverted Wishart distribution of course has some problems associated with it, as I discussed in the paper, and these problems relate to there being some inherent constraints imposed on the parameters within the distribution, which the analyst may find undesirable. This problem was first pointed out by Rothenberg, (1963). The argument is also summarized in Press, (1972, page 233). Nevertheless, the parameters of the inverted Wishart distribution may be assessed by assessing quantiles of the marginal distributions, which of course are inverted gamma distributions. The quantiles are related to variances, medians, etc. Methods for assesssing quantiles of univariate distributions are by now well known; see for example, Schlaifer, (1961); Stael von Holstein, (1970); Winkler, (1967a and 1967b); Lindley, Tversky and Brown, (1979). Methods for assessing the correlation or covariance for higher dimensional distributions are currently being developed; see for example, Gokale and Press, (1979); Dawid, Dickey and Kadane, (1979); Kadane, Dickey, Winkler, Smith and Peters, (1978). There are also several computer routines that have been developed to assist the analyst in assessing the hyperparameters of prior distribution families such as the inverted Wishart (see Press, 1980 for a summary). Methods for assessing the parameters of generalized distributions involving generalized hypergeometric functions have not yet been developed. Such methods will depend upon development of the theory that relates to these distributions in terms of marginal and conditional distributions. Once these procedures are known, methods that have already been developed can be readily applied.

Professor Dunsmore was surprised that the Bayesian development of QCF appeared much later than the earlier development. The explanation is of course, that the earlier development emphasized the use of qualitative controlled feedback as a data gathering tool, while the Bayesian development imposed some distributional structure above and beyond that which was assumed earlier, and this structure permitted us to make posterior inferences about results that might be obtained on a later round of QCF that we are not able to carry out. Such an analysis, while interesting and useful in some applications, is not as generally applicable as is the basic data collection process inself. In terms of practical relevance of the procedure it should be understood that the QCF approach can be easily implimented in a real world context for one, or even several, questions of importance without any application of the modeling itself. The practicality of the modeling stems from the fact that in our limited experience involving an empirical application of the methodology (see Press, Ali, and Yang, 1979) we found that after three stages, the process had pretty much stablized. We anticipate that only two or three stages will be necessary for stablization of the process in more general situations as well. Thus, if there were two stages, and we wanted to predict a third, and we used precisely the same numbers that Professor Dunsmore suggests in his comment, the dimension of the **a** vector would be 45*. It is of course always possible, and often reasonable, to keep the dimension of the **a** vector small by the device of using only those reasons for the prediction of the next stage's response, which were given by large numbers of respondents, and deleting the remainder. In that case, the dimension of the coefficient vector would always remain quite manageable. I will not comment further on the coefficients varying with the number of stages, beyond my saying that the model assumes that they do not so vary, in order to maintain a parsimonious approach to the number of parameters in the problem.

Professor Dunsmore talked about a respondent who might reverse his position from that on an earlier round, at some point in the process. This should occur only when some new information has been introduced into the composite of explanations for respondent's answers. If such a turn-about were not based upon new information, other group members would be totally confused and disappointed by the apparent lack of rationality of the turnabout group member.

In ignoring the variable with the "largest" variance matrix we are merely ignoring one observation out of many, and the ignored observation is one which is known with decreasing precision as the sample size gets large. Such an approximation can clearly have little effect on the result. I share the implied concern in Professor Dunsmore's final plea, which is to "see this elegant statistical theory transformed into useful statistical practice". It is my fond hope that practitioners of statistical methodology in various areas will apply qualitative controlled feedback to practical problems.

Professor French has made some excellent points. His first is that the methodology presented in this paper is applicable to a situation in which there is a single decision maker who plans to use the opinions of the group to help him make his decision. Thus, the decision maker is in fact exterior to the group. This point will be made later in my comments to Professor Mouchart.

Next I must talk about "old Charlie" who has a gut feeling for winning projects. I was not pursuaded by this argument because I don't agree with Professor French that, "some of the best opinions are held without the holder being able to express why he holds them". This is the same kind of argument used by anti-Bayesians to show why the entire Bayesian approach is not useful. They claim that while Bayesians must use prior distributions to develop their analyses, most people cannot really quantify their judgments, and for this reason, it is usually impossible to assess a prior distribution.

* Because the process stabilized, we could take $R_2(1) = R_2(2) = 0$. So if $R_1(1) = 4$, $R_1(2) = 5$, the dimension of the **a** vector would be 45. Moreover, it may often be assumed that $\mathbf{a}_\alpha(j)$ does not vary with α and j, in which case the number of distinct elements in **a** that must be estimated is $\pi(n-1)$; so if $\pi = 5$, $n = 3$, we must estimate a 10-dimensional vector.

The limitations of these types of arguments have been elucidated on many occasions, so I will not repeat them here. In the application described in this paper it is of course necessary for people to introspect about their opinions, just as they would regarding a prior distribution. In this case, they must introspect to derive arguments for why they believe what they believe.

Professor French's final point deals with the question of how representative are the results developed in a qualitative controlled feedback data collection process. The answer is that the results obtained after several stages of QCF are representative of what would be obtained if a census were taken of the entire population, and QCF procedures where applied. Thus, public policy or any other kind of policy, can be formulated for a large population based upon careful reasoning of a "representative" subgroup.

Professor Geisser raised the very interesting question, of whether or not the QCF procedure could be misused by an individual who was trying to control the outcome of the procedure? The answer is of course, that the procedure could in fact be misused by a devious intermediary. He could manipulate the outcome by misrepresenting the composite that was fed back to the panal on each stage. This is mentioned briefly in section one of the final form of the paper. It is not anticipated, however, that in most applications the context would be one in which manipulation is likely. Of course, the effect of manipulation can always be minimized by using a small group of intermediaries, rather than a single individual, to accomplish the task of forming the composite of reasons.

With respect to the issue of "What's it all about Alfie", there are simple and straightforward answers. Suppose, as Professor Geisser suggests, the "mayor of Valencia wanted to resolve a pressing public policy issue say the building of a Bayesian Conference Center". First, I would commend the mayor on his good taste, assuming he was the one who exercised the foresighted leadership to suggest such a center. Next, I would propose that he use qualitative controlled feedback on questions posed before a panel of people appropriate to the political context of Valencia (a city council, a random sample of concerned citizens, etc.) In an "open discussion", the Mayor (with the help of other high ranking, very local, and strongly influential people) might very well bully Valencia into a decision that is really inappropriate for this city. Using QCF the decision would have to be made through careful reasoning and rational dialogue. A popular referendum shares some of the features of careful reasoning with QCF, but because certain very vocal and affluent groups advertise heavily to persuade people to their position, regardless of the common good or the rationality of the argument, such decisions are often inappropriate. The social psychological literature abounds with examples of how special interest groups tend to dominate such "open discussions" (for a summary, see, e.g. Press, 1978).

Professor Mouchart asked about the properties of the opinion pooling process proposed in this paper? The answer to this question derives from the context in which this procedure should be evaluated. The context was carefully detailed and discussed in an earlier paper; see Press, (1978). There, it was pointed out that our context is one in which we always assume there is a single decision maker who wants to take every group member's opinions into account, but he will make the final decision. This is the same context assumed by Kirkwood, (1972), and it avoids the conflicts and difficulties addressed by the Arrow "impossibility theorem". As a result of using this context, conventional decision theory applies to any decision made by a decision maker on the basis of QCF.

I would like to close by thanking the individuals who where kind enough to comment on the paper in an effort to clarify the nature of the process being discussed. I am also particularly grateful to Professor Dunsmore for his thoughtful suggestions for improving the format of the paper.

REFERENCES IN THE DISCUSSION

ARROW, K.J. (1963). *Social Choice and Individual Values*, New York: Wiley.

DAWID, A.P., DICKEY, J.M. KADANE, J.B.. (1979). Distribution theory and assessment methods for matrix t and multivariate t models. *Tech. Rep.* University College of Wales, Aberystwyth.

DEGROOT, M.H. (1979). Comments on Lindley *et. al. J. Roy. Statist. Soc. A.* **142**, 172-173.

FRENCH, S. (1978). Consensus of opinion. *Euro. J. Opl. Res.* (in press).

GOKHALE, D.V. and PRESS, S.J. (1979). Assessment of a prior distribution for the correlation coefficient in a bivariate normal distribution. *Tech. Rep.* **58**, University of California, Riverside.

GOOD, I.J. (1952). Rational decisions. *J. Roy. Statist. Soc. B,* **14**, 107-114.

— (1953). The appropriate mathematical tools for describing and measuring uncertainty. Chapter 3 of *Uncertainty and Business Decisions,* 20-36. Liverpool: University Press.

— (1956). The surprise index for the multivariate normal distribution *Ann. Math. Statist.,* 1130-1135. Corrections, **1.** c. 28 (1957), 1055.

HARMAN, A.J. and PRESS, S.J. (1978). Assessing technological advancement using groups of experts. In *Formal Methods in Policy Formulation* (Bunn, D.W. and Thomas, H., eds.). 123-147. Basel: Birkhauser Verlag.

KADANE, J.B. and *et al* (1978). Interactive elicitation of opinion for a normal linear model. *Tech. Rep.* **150**, Carnegie-Mellon University.

KIRKWOOD, C.W. (1972). *Decision Analysis Incorporating Preferences of Groups.* Ph. D. Dissertation. Massachusets Institute University.

LINDLEY, D.V., TVERSKY, A. and BROWN, R.V. (1979). On the reconciliation of probability assessments (with discussion) *J. Roy. Statist. Soc. A* **142**, 146-180.

LUCE, R.D. and RAIFFA, H. (1958). *Games and Decisions.* New York: Wiley.

MURPHY, A.H. and WINKLER, R.L. (1977). Reliability of subjective probability forecasts of precipitation and temperature. *App. Statist.* **26**, 41-47.

PATTERNAIK, P.K. (1978). *Strategy and Group Choice.* Amsterdam: North-Holland.

PRESS, S.J. (1972). *Applied Multivariate Analysis.* New York: Holt, Rinehart and Winston, Inc.

— (1979). Qualitative controlled feedback for forming group judgments and making decisions. *J. Amer. Statist. Assoc.* **73**, 526-535.

— (1980). Bayesian computer programs. In *Bayesian Analysis in Econometrics and Statistics: Essays in Honor of Harold Jeffreys.* (Zellner, A., ed.) 429-442. Amsterdam: North Holland.

PRESS, S.J., ALI, M.W. and YANG, E. (1979). An empirical study of a new method for forming group judgments: Qualitative controlled feedback. *Technological Forecasting and Social Change*, **15**, 171-189.

ROTHENBERG, R.J. (1963). A Bayesian analysis of simultaneous equation systems. *Tech. Rep.* **6315** Rotterdam: Econometric Institute.

SCHLAIFER, R. (1961). *Introduction to Statistics for Business Decisions.* New York: McGraw-Hill.

WINKLER, R.L. (1967a). The assessment of prior distributions in Bayesian analysis. *J. Amer. Statist. Assoc.* **62**, 776-800.

— (1967b). The quantification of judgment: Some methodological suggestions. *J. Amer. Statist. Assoc.* **62**, 1105-1120.