DISCUSSION

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

Unlike Savage, I feel comfortable with the Bayesian position. I know of no case where it gives an unsatisfactory result and there are many cases where it produces an answer which is more satisfactory than others. This is not to say that there are no difficulties: there are, but they seem to be the sort that should yield to an adequate amount of research effort.

In the case of π, the resolution may lie in removing the excessive formalism sometimes imposed. There is a story that every paper appearing in the Annals of Mathematical Statistics had to have (X, A, P): a triple in which A is a σ-field of events. But why should we have a σ-field? In the case of π the σ-field is complicated and we could not do all the probability assessments demanded of it. What we do is to give *some* p-values, but not all. The important point is that those values given must be coherent. This, I believe, is essentially de Finetti's resolution.

Randomization is a puzzler to a Bayesian. Consider a trial to compare two medical treatments, T_1 and T_2. Suppose the result of the trial is that $p(\mathrm{R}\,|\,\mathrm{T}_1) > p(\mathrm{R}\,|\,\mathrm{T}_2)$ where R is the event of John's recovery, so that T_1 is preferred to T_2 for John. Suppose however there was a random quantity X such that $p(\mathrm{R}\,|\,\mathrm{X},T_1) < p(\mathrm{R}\,|\,\mathrm{X},T_2)$ for all X: would T_1 still be preferred for John? Such a set of inequalities constitutes Simpson's paradox. The paradox can be avoided if in the trial, X and the treatments are independent. Now a random allocation of treatments is, by the Bayesian meaning of random, independent of any X, so that a Bayesian might prefer randomization, though independence is what he is really after.

Rubin, (1978) shows that with randomization, the Bayesian calculations are much simpler. Simplicity is connected with the cost of rationality, mentioned by Savage. There has been little investigation of this and Savage does the conference a service by drawing our attention to the problem. When is it worth drawing a decision tree? Clearly the answer must depend on how well the utilities and probabilities can be determined. We usually draw the tree if they can: otherwise we might content ourselves with the initial utility. In the paper by myself, Tversky and Brown that Savage mentions, we studied the assessment of probabilities and suggested using measures of precision associated with them, rather like other determinations in science. Although this made the calculations complicated, and so more costly, my own feeling is that such measures are essential to any resolution of this important problem.

At this moment in statistics, my advice is to try the Bayesian paradigm. I think you will find that if works rather well.

The difficulties with Simpson's paradox also arise in considering the ideas put forward by Kadane and Sedransk. For suppose that the experts were all affected, either consciously or subconsciously, by X; then could it not happen that the resulting confounding of the treatment allocation with X would viciate the conclusions of the trial? Of course, if the confounding were recognized, it might be possible to allow for it: the real danger lies in an unrecognized confounding.

The criterion (1) has an attractive property: namely, it is invariant under linear transformations of the utilities and hence of the expected utilities. For suppose u_{tp} is replaced by $\ell_p u'_{tp} + m_p$, possibly a different transformation for each expert, then the

linear form $\Sigma u_{t_p} w_p$ becomes $\Sigma u'_{t_p}(\ell_p w_p) + \Sigma m_p w_p$ with the final term independent of p and the new weights $\ell_p w_p$ (it is not necessary that these add to one). I feel this is important since utilities are arbitrary up to linear transformations.

The authors admit the possibility that the patient may use his own utility function or have one supplied for him. May he not do the same for the probabilities? It is not clear to me that in evaluating my probabilities, I should use the expert's stated values, for I may feel him to be biased in some way. Thus if the motoring organization tells me their probability of getting stuck in the snow, I shall use a smaller value for my probability, since I believe they exaggerate the hazards in order to discourage people from using the roads and so reducing their chances of having to assist them in snowy conditions. As L.J. Savage pointed out, what the patient really needs is the expert's likelihood function (*not* his probability) to update the patient's prior.

A.M. SKENE *(University of Nottingham)*:

Professors Savage and Kadane though discussing quite different topics are both grappling with problems of utility. The first paper is essentially concerned with the utility structure of the decision problem "How shall I analyse this data?" while the second is concerned with the thorny problem of whose utilities to consider in clinica trials.

Professor Savage is concerned that the practical Bayesian doesn't practice what he preaches and claims that the decision theoretic framework doesn't take the cost of thinking into account. Now I believe I'm rational yet I would also guess the 29th digit of π and hope to win the bottle of sherry. It follows that in taking the decision to guess as opposed to computing the elusive digit I prefer the expected return from guessing viz 0.1 (Sherry - ϵ) + 0.9 (Nothing -ϵ) to the return (Sherry - Effort of computing - ϵ), where ϵ is the small effort necessary to make the snap judgement between the two alternatives. Thus it appears that I think that nine tenths of a bottle of sherry is not worth the extra effort. If I am happy with the outcome of this exercise in self enlightenment then my claim to rationality is unshaken for the time being at least; otherwise I must ponder afresh the fallibility of my decision making process.

There are two issues here. First, there is nothing in the decision theoretic framework which prevents one from including the cost of thinking. It can be incorporated quite naturally (in theory at least) as one of the attributes in a multi-attribute utility function. Secondly there may be some positive advantage in retrospection. It can be very much easier to see what a particular decision implies for the utility function, than attempting to assess the utility function directly. In medical decision making, for example, it is very difficult to get a Physician to assign utilities and costs to various treatments particularly when there is the possibility that a patient may be incorrectly allocated a treatment which is positively harmful. Having attempted such an exercise however, the Physician can then be observed making decisions on a long series of patients and from this information some idea of his actual utility function can be gained. (By considering which misclassification rates are considered aceptable for instance). The value of this exercise lies in the fact that reconciling the two utility functions so obtained may lead to better decisions in the future.

Consider a Statistician, invited to assist in the analysis of a set of data. He sees

several ways in which he might proceed and must choose one strategy. While the factors which influence the utility he has for each strategy are the personal choice of the decision maker, he might, for example, consider how far each strategy satisfied the experimenter's objectives and his own interests, the financial reward involved, the time necessary to execute each strategy, the computing cost/effort involved and perhaps how closely each strategy adhered to the principles of Bayesian statistics. In reaching a decision, the Statistician would, of course, find it necessary to choose weights reflecting the relative importance of these attributes. This situation is surely not unfamiliar. Perhaps we should be asking ourselves what weight we would give to the last of these attributes or, like the Physician, be thinking through the decision in abstract and then observing how we act in practice.

Turning now to the paper by Professor Kadane, I accept the author's remark that the paper is primarily a statement of intent; not a polished work where all the issues have been resolved, but rather an enunciation of a possible direction in which to proceed, together with the problems which are likely to be encountered.

In certain types of societies the practical solution to the ethical clinical trial problem could be achieved quite easily. When the allocation of a treatment depends on whose utility function you consider —the patient's or the physician's— we must combine the utilities in some way.

If the society is such that it sees it as the right or duty to define the role of an individual in a clinical trial then the problem vanishes. In the absence of such political involvement however the concept of an acceptable treatment seems an interesting one and worth investigating, though I am somewhat skeptical that these ideas will, in fact, lead to a new type of clinical trial.

A major problem as the author points out, is that in trials where there are no strong prior opinions it is possible for the procedure to converge to the wrong treatment and this leads the author to the idea of 'nearly acceptable' treatments. However, in practice no patients are denied treatment and in the standard randomised trial patients are all already receiving acceptable or nearly acceptable treatments. In effect, the current argument against randomised trials is based on the premise that 'nearly acceptable' is not 'ethical'.

It may sometimes happen that the patient's utility functions prevent certain treatment comparisons. Consider, for example, a trial comparing mastectomy with a form of radiation therapy for breast cancer. Given that all the participating physicians believe that there is little difference in efficacy and that effective treatment means survival as opposed to death then the utility functions of women involved will reflect preferences between the secondary consequences of the treatments, and thus, for example, the radiation therapy may be universally preferred.

Instances such as this of course don't prove that such a trial will never work. What is of greater concern is the possibility that such a trial is feasible but is misused. Will the utility functions of the participating experts be allowed to reflect things like loyalty to a particular company or the need to justify a particular research project to guarantee future funding? It is just conceivable that under the guise of an 'ethical clinical trial' more patients receive a less efficacious treatment than in a randomised trial. This would certainly be possible in trials where many patients were admitted to a trial before the

first results were known. Here, presumably, the experts could continue to use their prior opinions until the first results came to hand.

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

Professor Kadane points out that "an emphasis on the greater social good relative to the legitimate interests of the patient is less than satisfactory". I think we must distinguish between the patient's interests *before* and *after* he is known to be affected. For, it seems likely that, before he has got a particular disease, he maximizes his *personal* expected utility by voting a law which will oblige him to accept participate in a clinical trial were he to become ill and the trial necessary.

Indeed, he must balance his better chances of survival because of general scientific progress with the risk of having to accept a particular less efficacious therapy.

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

Since so much of the discussion of the various papers at this meeting has had a theological tone, it would seem appropriate to introduce the theological terms *probabiliorism* and *probabilism* to help describe the situation considered in this paper. In Webster's Third New International Dictionary we find the following definitions:

Probabiliorism - a theory that in moral questions where certainty is impossible only the more probable course may be followed.

Probabilism - a theory that in moral questions where certainty is impossible any course may be followed that is seen as solidly probable either through clear perception of the principles involved or through awareness of the support of judicious sound authority...any solidly probable course may be followed even though an opposed course is or appears to be more probable.

The authors seem to be urging us to be probabilists in carrying out clinical trials. The patient, however, must strongly hope that his doctor is a probabiliorist.

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

I have often wondered whether most clients who are given confidence intervals use them in some sense as Bayesian estimation intervals; see, for example, Good (1969, p. 184). Perhaps a sample survey is needed to answer the question at any moment in history, and for any field of application.

I have proposed a way of combining judgements of quantiles of distributions by various judges or experts in Good (1979) by methods rather different from those of Lindley, Tversky and Brown (1979). The application that directly provoked my work on this problem was the estimation of mineral resources. This application was brought to my attention by Dr. Larry S. Mayer.

To Professor Kadane, I have to say that, as a patient, I would be not happy with treatment A if two or more clinicians recommend treatment B and only one recommends treatment A, if I had no reason to prefere one clinician's judgment to those of others. I would prefer to accept a majority vote.

Lindley (1975) had an interesting idea for improving the ethics of medical trials,

but in Good (1978) his idea was shown not to be as applicable as it at first seemed. I said there that one way to make clinical trials ethical is to *pay* people to undergo them, and I doubt if this proposal was original, in fact it is already done when patients are given free treatment in exchange for entering the trial. Another way to pay patients, if they happen to be prisoners, is to give some remission of sentence as the form of payment. An objection that was raised in conversation by Dr. Kadane is that some are sentenced largely to keep them off the streets. To meet this objection the judge could be allowed to pass such sentences as: "Ten years without the right to enter medical trials and a further twelve years but with the right to enter such trials".

Another idea for making medical trials more ethical, when there is very little to choose between some treatments, would be to arrange to administer the treatments *simultaneously* to a sample of patients, perhaps at numerous medical centers. But this proposal might seldom be practicable.

A. O'HAGAN *(University of Warwick)*:

All statistics in practice is approximate. Perfect analysis requires an infinite amount of effort to achieve. Therefore the ideal of rationality must be tempered by pragmatism. This theme lies at the back of several papers at this meeting but I think Professor Savage hits the nail on the head when he relates it to the cost of effort. The degree of approximation finally accepted in any analysis results from a balance between the gains accruing from more nearly optimal decisions which might be made with an improved approximation, and the cost of that improvement. The cost of better approximation may have many components, but the costs of thought and of computer time spring quickly to mind. Professor Savage suggests that it may be possible to measure these costs, but I doubt if that would help much because the measurement of the gain from improved approximation is much more difficult. It obviously depends on the true analysis, which is unknown, and any attempt to theorise about it will introduce new quantities which themselves must be approximated in practice. Statistics will always be a matter of subjective, unformulated judgements. As Professor Good says in his paper, "I stop when the guessed expected utility of going further becomes negative if the cost is taken into account".

The fact that no practical statistics can ever be more than an approximation to the ideal Bayesian analysis is no reason to despise Bayesian principles and theory -that is the trap into which Dr. Leonard nearly falls with his paper at this meeting. Theory serves at least two distinct purposes. First it provides guidelines. If we know that a certain analysis is optimal for a given problem which we can think of as approximating our own problem, then that analysis serves as an initial approximation for us. Some thought about ways in which the real problem deviates from the theoretical one suggests (by reference to other theory) ways in which we should modify our initital analysis. Dr. Leonard acknowledges this role of theory but gives the impression that it is unimportant, yet without the guidance of theory the applied statistician would be completely lost.

The second purpose is to reduce costs. A new piece of theory means that in appropriate circumstances the statistician can proceed inmediately to a greater accuracy of approximation with only slight costs in extra thought or computing. Professor Sava-

ge recognises this, particularly in section 4. The investment that the theoretician's thought represents can yield rich dividends for the practitioner.

REPLY TO THE DISCUSSION

I.R. SAVAGE (*Yale University*):

As noted earlier I claim no originality for this essay. Good's (1979) review of Colodny (1977), hints at my borrowing from L.J. Savage as expressed in his late essay "The Shifting Foundations of Statistics".

This Conference's success should be evaluated in terms of its helping to create theory and application of Bayesian statistics. In doing this there is no last word. I am glad I had the opportunity to participate and I'm thankful for the lively remarks of the discussants.

J.B. KADANE (*Carnegie Mellon University*) and N. SEDRANSK (*S.U.N.Y. at Albany*):

We thank Professors Bernardo, DeGroot, Good, Lindley and Skene for their attention to the problems we pose and for their useful ideas.

The point made by Professor Skene and also raised by Professor Bernardo, that societies differ in the degree of coercion they exert on their members, is unarguable. Hence, the balance between the rights and interests of ill citizens seeking the "best possible" treatment and the rights and interests of the rest of the society in fostering medical research is not uniquely defined for all nations and societies. In the United States, it has been required for some time that a patient consent to participation in a clinical trial prior to the beginning of any therapy or procedure under study, and further that the patient's consent must be given with full knowledge of all relevant information currently available. In this context, both the legality and the feasibility of a clinical trial revolve about the question of whether or not a patient rationally would give informed consent to participate in the clinical trial. Thus the statistical design and analysis must address this question and a Bayesian approach provides a natural formulation. (Consideration of optimal legislation to define a new context for human experimentation is beyond the scope of this paper).

The intent of this paper is to formulate a model for clinical trials which would embody both the rights of the patient and the rights of society and which would exploit the variety of expert opinions within the scientific community to justify study of alternative therapies. Both Professors DeGroot and Lindley express concern that the patient be allowed to modify the expert opinions and/or reject selected expert opinions altogether. When a patient considers entering a clinical trial, he acquires a collection of experts beyond the particular physician he consults directly. A very sophisticated patient might want to correct for the several physicians' varied biased; this modification presents fewer mathematical difficulties than practical obstacles. A much less sophisticated patient might choose to ignore all opinions except that of the physician he consults. However, this model for a clinical trial returns to that of an "uncontrolled" trial, which offers less assurance that the best treatment will be identified correctly.

Sources of potential vulnerability of this class of designs are viewed with concern by both Professors Lindley and Skene; but whereas Professor Lindley considers the possibility of unrecognized factors, Professor Skene worries over infelicitous configu-

rations of known factors. The confounding of unrecognized factors with treatment effects can vitiate the results of any study. To the extent that an unrecognized factor is correlated with a recognized factor, its effect is controlled by the incorporation of the recognized factor in both the statistical design and the statistical analysis. Of course, if the unrecognized factor is independent of all the recognized factors, its occurrence in a pattern resulting in confounding its effect with the treatment effect requires two similar sequences for the sequence of treatment assignments and for the sequence of values for unrecognized factor. (Then purely probabalistic arguments apply; and, for example, in the case of an unrecognized binary factor, the risk of confounding is minimized for balanced designs). Untoward influence of recognized factors, can, as Professor Lindley points out, be averted by proper design and analysis. Precisely for this reason, it has been presumed throughout the paper that covariates are used directly in the probability distributions, and therefore are included in both the design and the analysis.

It is certainly possible, as Professor Lindley suggests, that a physician could inject his (knowing or unwitting) bias into the clinical trial by way of his prior distribution. One strength of the class of designs proposed is its *lack* of vulnerability to a single physician's bias. Only in the event of a rather uniform bias on the part of all experts should the confounding occur, a somewhat less likely possibility than the occurrence of a significant bias on the part of a single physician.

Professor Skene's concerns about failures of the clinical trial design caused by known factors, pose much smaller problems since these possibilities can be examined specifically for each trial before it starts. The possibility that the patients' utility functions might prevent certain treatment comparisons is not well illustrated; in fact, in the circumstances Professor Skene cites, no patient rationally would agree to a randomized trial. Professor Skene also expresses apprehension that less than honest scientists could exploit this class of designs. The use of fraudulent experts can ruin a clinical trial of almost any design; in the class of designs proposed here, as in all designs for scientifically responsible research, experts with conflicts of interests are assumed to be ineligible for influential decision-making roles. The possibility that there is insufficient agreement among experts or that there is near unanimity preventing either initiation or termination of a clinical trial can best be examined by simulation; comment must be deferred until these simulations have been completed.

Professor Good's notion that a trial might be made ethical by simultaneous administration of treatment at several medical centers bears some resemblance to what is now done. But doesn't it address the ethical issue by failing to generate relevant data for any of the patients? A more sequential approach would yield early returns and might avoid giving bad treatments to at least some of the patients. Payment for prisoners in direct form or by sentence reduction is specifically prohibited within the U.S.; and remuneration to non-prision participants may not be of such a magnitude as to impair the individual's judgment of the medical merits of the options offered him.

All these issues are difficult, and well worth discussion and further research. We are grateful to our discussants for their stimulating thoughts.

REFERENCES IN THE DISCUSSION

COLODNY, R.G. (ed.) (1977). *Logic, Laws and Life: Some philosophical considerations.* Pittsburg: University Press.

GOOD, I.J. (1969). What is the use of a distribution?. In *Multivariate Analysis-II* (P.R. Krishnaiah, ed.) 183-203. New York: Academic Press.

— (1978). Ethical treatments. In *J. Statist. Comput. and Simulation* **7**, 292-295.

— (1979). On the combination of judgements concerning quantiles of a distribution with potential application to the estimation of mineral resources. *J. Statist. Comput. Simulation* **9**, 77-79.

— (1979). Review of Colodny (1977). *J. Amer. Statist. Assoc.* **74**, 501-502.

LINDLEY, D.V. (1975). The effect of ethical design considerations on statistical analysis. *Applied Statistics* **24**, 218-228.

RUBIN, D.B. (1978). Bayesian Inference for causal effects: the role of randomization. *Ann. Statist.* **6**, 34-58.