What both frequentists and Bayesians tend to overlook is the fact that data (and data summaries) are put to many uses and viewed from many perspectives (prior opinions) by a statistician's audience. Concentration upon which single procedure to use in analyzing data tends to neglect the diverse interests of this audience. More attention needs to be paid to the design of the experiment, The founders (Neyman, Pearson, Wald) of the modern frequentist approach to inference were aware of this point, and discussed designs based on minimax procedures as a way to satisfy all users of the data. Where prior opinion is not highly variable (as assumed by the minimax approach), designs constructed from a robust Bayesian perspective may be more efficient in satisfying the needs and interests of a statistician's audience. A step in this direction has been made in hypothesis testing by my student Burt (1989), and for estimation by DasGupta and Studden (1989).

REFERENCES

BURT, J. (1989). Towards Agreement: Bayesian Experimental Design. Ph.D. dissertation, Dept. of Statist., Purdue Univ.

DASGUPTA, A. and STUDDEN, W. J. (1989). Towards a theory of compromise designs: Frequentist, Bayes and robust Bayes. Mimeo Series No. 89-20, Dept. of Statist., Purdue Univ.

GLESER, L. J. (1987a). Improved estimators of mean response in simulation when control variates are used. Mimeo Series No. 87-32, Dept. of Statist., Purdue Univ.

GLESER, L. J. (1987b). Improved estimators of mean response in simulation when control variates with unknown covariance are used. Mimeo Series No. 87-34, Dept. of Statist., Purdue Univ.

JOHNSTONE, I. (1987). On inadmissibility of some unbiased estimates of loss. In *Statistical Decision Theory and Related Topics IV* (S. S. Gupta and J. O. Berger, eds.) **1** 361–379. Springer, New York.

LU, K. L. and BERGER, J. O. (1989). Estimation of normal means: Frequentist estimation of loss. *Ann. Statist.* **17** 890–906.

DEPARTMENT OF MATHEMATICS
AND STATISTICS
UNIVERSITY OF PITTSBURGH
PITTSBURGH, PENNSYLVANIA 15260

B. M. HILL

The University of Michigan

I am delighted for the opportunity to discuss this interesting paper by Professor Brown. I have long admired his work both for its technical virtuosity and for his thoughtful, philosophical discussions of the statistical approach he advocates. I have often expressed the hope that frequentists would offer Bayesians some interesting challenges, which Professor Brown does here. He shows that the frequentist admissibility paradigm (FA) is not compatible with the principle of ancillarity (AP) and suggests that the latter should be abandoned. AP is an intuitively compelling idea, a version of which is implied by the restricted likelihood principle of Hill (1988), which in turn is implied by the subjective Bayesian approach that I advocate. Therefore there is a clear conflict between basic principles, which must be analysed carefully, to see which, if any, can hold up. In the history of science it is such conflicts that often lead to rapid progress.

I shall try to argue that it is FA in the context of infinite partitions that cannot be maintained. My criterion, which must be an "objective" one in order for others to accept my conclusions, will be totally with regard to which of the two principles is most relevant for the practice of statistics, whether inferential, predictive or decision-making. Without such a criterion for choice, the question would be merely academic and would reduce to a question of taste. FA will be defined by Professor Brown himself in his discussion of an article by Berger (1984, page 126):

> My perspective can be described in a few words as that of a pure, collective, pragmatic frequentist...
> Here is the basic principle: Statistical (and other) procedures being used today and being proposed for future use should be judged collectively and realistically according to their long term expected consequences.... .
> The preceding is a statement of general philosophy, not a mathematical axiom. This principle is to some degree connected with an acceptance of fundamental axioms of probability theory such as those of Kolmogorov.... .

I regard the spirit of FA as admirable. This spirit underlies all attempts to eliminate the gibberish from statistics and decision-making. This is especially important nowadays, when statistical methods are more and more being applied to important real-world problems, such as environmental matters, assessment of health care and public policy in general. Historically, there has been an unfortunate conflict between two quite distinct points of view: one the view of a creative scientist, such as R. A. Fisher; the other the view of a policy-maker or perhaps even a bureaucrat. The former is primarily concerned with unique instances of scientific activity, and not with the results of a procedure that is repeated many times. For example, in considering the evidence concerning the greenhouse effect or pollution of the atmosphere, it does not seem to be of much relevance to consider hypothetical repetitions of the overall experiment. One wants instead to understand the particular phenomenon in question. Data analysis is of greatest interest in this sort of problem and can be coupled with the Bayesian approach to deal with inference and decision-making, as discussed in Hill (1990). On the other hand, the policy-maker view primarily concerns the consequences of repeated use of a certain policy. There is no conflict, in my opinion, between these two points of view. Both are of great importance. I view the latter, however, as the one that is of primary interest in connection with Professor Brown's example, and it is only this that will be discussed here. The thesis I will maintain is that even for the policy-maker problem, the FA principle, as employed by Professor Brown, is not appropriate.

I would like to begin by discussing the Stein paradox. For a time this paradox confused me, since it too involved a conflict between the FA principle and the subjective Bayesian approach. In the late 1960's I began to ask questions about the operational meaning of inadmissibility, and to my surprise, could find not even a single attempt *operationally* to justify the admissibility principle. It was simply taken as self-evident that inadmissible estimators should not be used. It is well known in the history of science that it is precisely such implicit assumptions that often prove untenable. Examples are the notions of absolute space and time of Newton, or of simultaneity of events, prior to the special theory of relativity. Eventually I put forth an argument in Hill (1974, Section 2) that purported to provide an objective sense in which one might argue against inadmissible procedures. The analysis is brief, and I would like to repeat it here, since I think that the application of this argument is at the core of the issue.

Let θ be a conventional parameter that determines the distribution of a random variable X and let $d_i(X)$, $i = 1, 2$, be two decision functions that depend on the value of X. Suppose that we are in a situation that is repetitive and that $L(\theta, d_i(X_j))$ is the loss to be sustained if d_i is used on the jth occasion, $j = 1, \dots, M$. Then the expected losses, given the value of θ, are $R_i(\theta) = E_{X|\theta} L(\theta, d_i(X))$, $i = 1, 2$, which are the usual risk functions. Here X is a generic random variable having the same distribution as each X_j. [There are in fact some subtleties with regard to this framework of repetitive situations, due to the possibility of learning from one occasion to another. See Hill (1974, page 560). However, although oversimplified, I regard the basic scenario here as valid for the policy-maker problem. For example, one can suppose that the decision functions are mechanically implemented on a computer, without the statistician actually observing the X_j, so that no learning can take place from one occasion to another.] Suppose now that there is a referee who generates couples (θ_j, X_j) on a computer, for $j = 1, \dots, M$, using a probability distribution π to generate the θ_j, and some conditional distribution for X_j, given θ_j. Let the referee generate M independent couples in this way. Assume that the conditional distribution for X_j, given θ_j, is known to all concerned, but not π. In this case the loss associated with use of d_i on the jth occasion is $L(\theta_j, d_i(X_j))$. Summing over the M occasions, the actual increment in loss if d_2 were used on each occasion instead of d_1, would be $\sum_{j=1}^{M}[L(\theta_j, d_2(X_j)) - L(\theta_j, d_1(X_j))]$. The expectation of this increment from the perspective of the referee who knows π is then $K \times M$, where K is the π expectation of $K(\theta) = R_2(\theta) - R_1(\theta)$. If d_2 is dominated by d_1 in the sense of inadmissibility, then $R_2(\theta) - R_1(\theta) \geq 0$, with the inequality strict for some θ. So if π gives positive weight to the set of θ where the inequality is strict, then $K > 0$. This evaluation is mathematically valid provided only that Fubini's theorem holds, and so would be true for loss functions bounded from below.

If π has positive mass where $R_2(\theta) - R_1(\theta) > 0$, then I argued that from the perspective of the referee, who knows π, it would be imprudent to use d_2 in preference to d_1, and referred to any law pertaining to large M (laws of large numbers, etc.) under which the overall loss could be regarded as likely to be large if M were large. The reference to a referee is made primarily to represent the situation where there is some mechanism which generates what might be called a "true" distribution for θ, so that one can assess the performance of the two decision functions from the perspective of such a distinguished distribution. It is not necessary that such a referee exist, but only the mechanism. Note that if the inequality $R_2(\theta) - R_1(\theta) \geq 0$ were strict only at one isolated point, for example some undistinguished irrational number, then the argument would apply in its weakest form and K would be positive only if π were to give positive probability to this point. On the other hand, if d_2 were inadmissible in the extended sense, i.e., could be dominated uniformly for some $\varepsilon > 0$ and was in this sense dominated by d_1, then for any π it would follow that the increment in expected loss would be at least $\varepsilon \times M$. This suggests that even within the conventional decision theory framework, admissibility in the extended sense might be a better criterion than admissibility, since the above evaluation would then not depend upon the choice of π. The inadmissibility that occurs with respect to least squares estimators is intermediate between these two extreme cases, since the risk function for the least squares estimator can be improved upon everywhere, but not uniformly by some fixed ε. See Heath and Sudderth (1978), who prove that in the scenario of the Stein paradox, least squares estimators are admissible in the extended sense. Although not entirely convincing, since in many examples neither a referee nor a mechanism are known to exist, to the best of my knowledge this remains essentially the *only* argument that has ever been given operationally to justify admissibility as a criterion. The same basic argument is given by Berger (1984, page 88; 1985, page 257).

I now believe that even this argument requires some qualification in order to be truly operationally meaningful. It must be noted that a computer has only a finite memory and in a finite time can produce numbers (or vectors) only within a certain domain, which can be taken as a grid of equally spaced points bounded in magnitude by a known constant. I would like to argue that any supposedly bad consequences to be associated with the use of a statistical procedure should be required to be demonstrable in the context of a game with prespecified rules, that can with certainty be played and completed by two players in a finite time, with a referee in charge to keep order, and on a computer with a known finite bound on its memory. (We need not know the precise memory, but merely have an upper bound, which I suppose to be known to both players.) This in effect will constitute my definition of "operationally meaningful". It should be noted that although our conventional statistical models often involve probability densities on the real line, all observable and recordable variables must necessarily be rational valued. No irrational number can be recorded in a finite time, nor can it be demonstrated by measurement that any empirical quantity is irrational.

I believe that the above qualification of my original argument is important. It is well known that in the finite case, if a procedure is admissible then it is a Bayes procedure, and if it is Bayes for a prior distribution with all probabilities positive, then it is admissible. See Blackwell and Girshick (1954, page 127) and DeGroot (1970, page 133). In fact, when the parameter partition is a finite one, the admissible rules are necessarily contained in the closure of the class of Bayes rules for prior distributions with all probabilities positive, and so the only mathematical distinction between the class of admissible rules and the class of such Bayes rules is that between a set and its closure. For a computer with finite memory, so that computable quantities are necessarily finite both in number and in number of digits, such a distinction would not appear to be of great importance. So I would argue that the conflict between FA and AP that Professor Brown introduces *cannot occur in the case of parameters and data for which there are only a finite number of possibilities, and this is the only situation that can be implemented on a computer with finite memory, in finite time and according to prespecified rules.*

I do not think it is generally perceived just how subtle are the considerations that arise when attempting to use infinite models operationally. The conventional use of Euclidean spaces and probability densities with respect to Lebesgue measure in mathematical statistics is based primarily on the mathematician's desire for generality. It is easy to imagine that because the rational numbers are all contained in the real line, that therefore there is no possible harm in making an extension to the more general case. But I think this is false, and I will now state and prove a simple theorem to emphasize the point. Let X be an inadmissible estimator in the usual context, i.e., with the parameter space some Euclidean space, and let Y be an estimator that dominates X in the sense of inadmissibility. Suppose that X is Bayes with respect to some improper prior distribution on its parameter, which for convenience we shall take to be a uniform improper density, i.e., Lebesgue measure. (For example, X might be the least squares estimator proved inadmissible by Stein in dimension at least 3, and Y might be the corresponding James–Stein estimator.) Let $\mathscr{P}$ be any fixed known finite grid of points in the parameter space, for example, points with coordinates that are multiples of some fixed tiny rational number. Let $X_{\mathscr{P}}$ and $Y_{\mathscr{P}}$ be the corresponding estimators as implemented on some fixed computer, with all numbers being rounded to some prespecified number of decimal points. Since X itself is Bayes with respect to a uniform prior distribution, let $X_{\mathscr{P}}$ be defined as the Bayes estimator for a uniform prior distribution on the points of $\mathscr{P}$. $Y_{\mathscr{P}}$ can be obtained by any prespecified algorithm for computing Y. Note that because of the implementation on a computer with finite memory, X must differ from $X_{\mathscr{P}}$ and Y must differ from $Y_{\mathscr{P}}$. By *admissible*$^{\mathscr{P}}$ we shall mean admissibility with respect to the parameter space restricted to the finite grid of points represented by $\mathscr{P}$.

THEOREM 1. *For each $\mathscr{P}$, $X_{\mathscr{P}}$ is admissible$^{\mathscr{P}}$. For each $\mathscr{P}$, $Y_{\mathscr{P}}$ is either itself a Bayes estimator with respect to $\mathscr{P}$, or else is not admissible$^{\mathscr{P}}$.*

PROOF. For every $\mathscr{P}$, $X_{\mathscr{P}}$ is Bayes against a uniform prior on this finite grid and is therefore admissible$^{\mathscr{P}}$. For finite partitions of parameter and data, the admissible class is contained in the Bayes class, so if $Y_{\mathscr{P}}$ is not Bayes for $\mathscr{P}$, then it cannot be admissible$^{\mathscr{P}}$. □

This theorem implies that if one takes FA with the usual infinite partition for the parameter space seriously, and therefore chooses Y in preference to X, then one is implicitly rejecting FA for *every* finite grid of points. In fact, one is choosing the (typically) inadmissible implementation $Y_{\mathscr{P}}$ in preference to the admissible implementation $X_{\mathscr{P}}$! But it would instead appear to be in the spirit of the FA paradigm to choose the *implementation* that has desirable risk properties, rather than to be concerned with the way in which the rules are derived.[1] This is why I do not think it is so innocuous to adopt such infinite parameter spaces. It should be remarked that an estimator that is known to be inadmissible, such as the James–Stein estimator, may turn out to be such that its implementation $Y_{\mathscr{P}}$ is *admissible*$^{\mathscr{P}}$ for some $\mathscr{P}$. This is not ordinarily to be expected, but it can occur. In this case, the choice between $X_{\mathscr{P}}$ and $Y_{\mathscr{P}}$ would be between two Bayes estimators, each of which is *admissible*$^{\mathscr{P}}$. It would then be helpful to know for which prior distribution $Y_{\mathscr{P}}$ is Bayes. There is a lot to be said, on the grounds of objectivity and custom, in favor of $X_{\mathscr{P}}$, unless the a priori distribution for which $Y_{\mathscr{P}}$ is Bayes can be given some compelling motivation. The James–Stein estimator can sometimes be given some motivation via the Bayesian analysis of random effects models, as in Lindley and Smith (1972) and Hill (1974, 1977, 1980b).

The upshot of the theorem is that it may be more important than is generally realized to think through the subtle issues involved in the use of conventional infinite parameter spaces. Thus from the theorem it follows that any supposed inadequacy or paradox concerning the least squares estimator must ultimately depend upon an inadequacy in representing the real-world situation by *any* finite partition. In choosing to take FA seriously for infinite partitions of the parameter, one is implicitly ignoring it entirely for *all* finite representations. It should be remarked that even apart from the fact that computers have finite memory and that recordable data is necessarily rounded, there are still other reasons to consider the finite case as the most meaningful. Ordinary measurements cannot be indefinitely refined, for one thing because the quantity being measured usually does not have meaning beyond a certain point and may vary with time. Human height is known to vary during the course of a day. Even time itself is considered to be discrete by some modern physicists [Whitrow (1980), page 203]. Although it would appear impossible ever to entirely rule out the infinite case, at least in certain exotic scenarios arising in the physical sciences, these would hardly be typical of applications in the biological and social sciences or in real-world decision-making. Finally, if in some example one does take seriously the possibility that the parameter does

[1] In fact this would even be true for the implementation of $X_{\mathscr{P}}$.

not lie in any finite grid, then one can consider the gambles to be conditional gambles, conditional on the parameter being in some finite set. In this case the gambles or losses are called off if the parameter is not in the set. See my discussion of Berger and Wolpert (1988, page 167) where I argue that the extended inadmissibility that arises in the Monette–Fraser example has no operationally bad consequences in the case of a computer with finite memory and such conditional gambles.

There are in fact infinitely many paradoxes of the infinite. What constitutes a paradox is itself a matter of taste. What some would describe as paradoxical, seems quite trivial to others. Often it is the better mathematician–philosopher who sees through supposed paradoxes, but occasionally it is otherwise. For example, after a few years of calculus, some think that the ancient paradoxes of Zeno are trivially resolvable, while a mathematician–philosopher of the rank of Russell (1937, page 347) has instead argued that some of the paradoxes of Zeno have never been fully resolved. He describes the paradoxes of motion as "all immeasurably subtle and profound." Similarly, one can lose oneself forever in various paradoxes of the infinite, such as the antinomies of Kant, about which there is still no agreement despite two centuries of analysis. Note also that even supposing that we had a computer with an infinite memory, what would it mean to draw a random sample from an infinite set? Some mathematicians and logicians question the relevance of the axiom of choice in proving mathematical theorems. [The intuitionist school of mathematics does not accept the validity of the law of the excluded middle in dealing with infinite sets. This school is constructivist and it is in sharp contrast to the formalistic school of mathematics, in which "... much or all of pure mathematics is a meaningless game." Hersh (1979), page 12.] Thus, given that such paradoxes are so abundant, and so unresolvable, I do not think that a result that depends upon hypothetical operations with infinite sets, can be found entirely convincing by the statistician who must analyse data on a computer with a finite memory or by the decision-maker who must utilize the results of such a statistical analysis.

On the other hand, if we take the finitistic view that I advocate here, everything is operationally meaningful, and it seems to me that we have a clear cut argument for use of FA in this context. My conclusion from this discussion is that the paradox of Professor Brown is only meaningful if one accepts the appropriateness of FA in the context of an infinite partition of the parameter, and to me this means it has little relevance to real-world problems. On the other hand, AP is compelling even for finite partitions, as for example, in the example of Cox, which can be modified so as to apply in a finite scenario. So logically speaking, Professor Brown is asking us to reject a principle (AP) for the case of practical interest, the case of *some* finite representation of the parameter space and data, because of a mathematical result that cannot arise in precisely this case.

I do not mean to be too hard-nosed on the finite–infinite question, because it seems to me that infinite models do play a very valuable role in providing approximations and insight, and indeed were developed by great mathematicians such as Laplace, Gauss, Poincaré and others, for precisely this purpose. But here one must be very cautious with regard to the precise sense and relevance of each particular approximation that is being made. I do not think anyone regards it as paradoxical that an approximation that works well in one situation works poorly in another. It is this that I believe has been lost in the conventional approach to decision theory, where a passage to the limiting partition for the parameter is made without even an attempt at justification and without due regard to the sense in which such a passage provides an approximation to the case of real interest, the finite case. Thus my question for Professor Brown is whether he can provide an operationally meaningful argument for admissibility as a criterion in the case of an infinite partition, in such a way that it is clear that someone who used an inadmissible procedure would suffer an objectively verifiable loss? If Professor Brown can show how this can be done, it would be a valuable contribution. But even then, one would be perfectly safe in all applications that do involve only a finite scenario.

Next I would like to discuss countable additivity in much the same terms. The distinguished probabilists who founded and developed the modern measure–theoretic approach to probability were under no illusions as to the truth or validity of the assumption of countable additivity or its equivalent, the axiom of continuity. For example, Kolmogorov (1950, page 15) says:

> For infinite fields, on the other hand, the Axiom of Continuity, VI, proved to be independent of Axioms I–V. Since the new axiom is essential for infinite fields of probability only, it is almost impossible to elucidate its empirical meaning, as has been done, for example, in the case of Axioms I–V in 2 of the first chapter. For, in describing any observable random process we can obtain only finite fields of probability. Infinite fields of probability occur only as idealized models of real random processes. *We limit ourselves, arbitrarily, to only those models which satisfy Axiom VI.*[2] This limitation has been found expedient in researches of the most diverse sort.

Now while expediency may be temporarily of importance, it can hardly be considered in the nature of fundamental truth. At a certain time in history it was convenient to use the abacus and at another time the slide rule, but now for most of us it is not. If we look into the axiomatic approaches to probability and statistics of de Finetti (1974) and of Savage (1972), we see that the former has been highly critical of the axiom of countable additivity, and his criticisms have never been answered, while Savage's fully rigorous theory was not able to establish countable additivity. Now I believe that Kolmogorov and other great probabilists, such as W. Feller, while knowing that countable additivity cannot be justified, were under the impression that it was not likely to do much harm.

[2]Author's italics.

But as de Finetti (1974, page 33) has argued, this is not true, since it forces one into a "Procrustean bed" in which in effect certain types of approximation are eliminated. Ramakrishnan and Sudderth (1988) have shown that even in the simplest of all probability scenarios, that of flipping a fair coin, Borel's strong law does not hold in the finitely additive context. These authors show that with exactly the same joint distributions for all finite sequences, i.e., probability $1/2^k$ for any k-tuple of 0's and 1's, one can have the average converge everywhere to 0, converge everywhere to 1 or fail to converge everywhere. This means that *no finite experience* with a coin can possibly determine what happens in the limit as the number of trials goes to infinity. The usual convergence in the conventional theory occurs simply because it is contained in the assumption of countable additivity, and the argument is circular.

This is one reason why my argument for admissibility focuses on only a finite number of repetitions M. No one can really say anything about what happens as $M \to \infty$ other than in terms of the weak law of large numbers or various approximations, such as the central limit theorem. This can be viewed again as saying that one cannot simply pretend that M is very large, as though that was all there is to say. The approximation that one obtains in this way is simply not good enough, and it becomes even worse in the context of two-dimensional data, such as in the Borel, Dubins and Stone paradoxes discussed in Hill (1980a). Here we have the paradox of nonconglomerability, i.e., where a probability need not lie between the inf and sup of its values given the elements of a partition, of which de Finetti (1974) has given many examples. I do not now believe even nonconglomerability to be truly paradoxical, although it is much more surprising than inadmissibility. In nonconglomerable situations, one can be faced with sure loss of at least some fixed $\varepsilon > 0$, given each value of a parameter, say θ. Such sure loss implies both inadmissibility and extended inadmissibility. See my discussion of Berger and Wolpert (1988, page 164) for the connection with the Stein paradox. I once thought that nonconglomerability might possibly be paradoxical, but now think that the paradoxical aspect resides instead in thinking that the use of infinite partitions has operational meaning. Schervish, Seidenfeld and Kadane (1984) and Hill and Lane (1985) show that for countable spaces, countable additivity is equivalent to conglomerability. See also Scozzafava (1984).

Still another reason to reject the notion of indefinite repetition in the argument for admissibility given above is because of the fact that conditions in this world appear constantly in flux. A wise ancient Greek once said that one can never cross the same river twice, to which another responded, "Not even once." Thus all things change, which has some serious implications for the notion of an infinite i.i.d. sequence of random variables, and taking too seriously the representation of parameter values as infinite decimals.

The most basic reason that I reject the notion of countable additivity and other related ideas is because in my opinion all probabilistic and statistical models should at the very best be regarded only as reasonable approximations. Such models can be very useful, but must *always* be taken with a substantial grain of salt. To take literally a model such as the Gaussian, which in theory produces *irrational* numbers (none of which could ever be recorded even if there were such a process), causes great confusion, and even the most careful experiments, such as the celebrated Michelson–Morley experiment of physics, will introduce bias that will overwhelm other considerations, such as to how many decimal points the numerical value of a parameter has meaning. I do not, however, object to the careful use of standard parametric models in scientific research. Such models are often approximately valid and are useful both for communication and insight. The model of Professor Brown is a generalization of the important random effects model, and as an approximation has numerous real-world applications. But it is one thing to offer valid alternatives to least squares estimators, for example, based upon the random effects model, and another thing to suggest that least squares estimators are in some sense objectively inadequate because of a criterion such as admissibility that has been translated to the idealized world of infinite precision of both data and parameter values.

In conclusion, statistical decision theory, insofar as it pertains to public policy and other human affairs, will inevitably be implemented on computers with finite memories, using finite representations for the data and parameters. The Bayesian approach for finite partitions works very well, as I am sure Professor Brown would agree. If he can extend the FA argument to the infinite case in a way with clear operational meaning this would be very valuable. Until this is done I do not think it is warranted to impugn the validity of either AP or least squares estimators.

REFERENCES

BERGER, J. (1984). The robust Bayesian viewpoint (with discussion). In *Robustness of Bayesian Analysis* (J. Kadane, ed.) 321–372. North-Holland: Amsterdam.

BERGER, J. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. Springer, New York.

BERGER, J. and WOLPERT, R. (1988). *The Likelihood Principle*, 2nd ed. IMS, Hayward, Calif.

BLACKWELL, D. and GIRSHICK, M. A. (1954). *Theory of Games and Statistical Decisions*. Wiley, New York.

DE FINETTI, B. (1974). *Theory of Probability* **1**. Wiley, London.

DEGROOT, M. (1970). *Optimal Statistical Decisions*. McGraw-Hill, New York.

HEATH, D. and SUDDERTH, W. (1978). On finitely additive priors, coherence, and extended admissibility. *Ann. Statist.* **6** 333–345.

HERSH, R. (1979). Some proposals for reviving the philosophy of mathematics. In *New Directions in the Philosophy of Mathematics* (T. Týmoczko, ed.) 9–28. Birkhäuser, Boston.

HILL, B. M. (1974). On coherence, inadmissibility and inference about many parameters in the theory of least squares. In *Studies in Bayesian Econometrics and Statistics in Honor of L. J. Savage* (S. Fienberg and A. Zellner, eds.) 555–584. North-Holland, Amsterdam.

HILL, B. M. (1977). Exact and approximate Bayesian solutions for inference about variance components and multivariate inadmissibility. In *New Developments in the Application of Bayesian Methods* (A. Aykac and C. Brumat, eds.) 129–152. North Holland, Amsterdam.

HILL, B. M. (1980a). On finite additivity, non-conglomerability, and statistical paradoxes (with discussion). In *Bayesian Statistics* (J. M. Bernardo, M. H. Degroot, D. V. Lindley and A. F. M. Smith, eds.) 39–66. University Press, Valencia, Spain.

HILL, B. M. (1980b). Robust analysis of the random model and weighted least squares regression. In *Evaluation of Econometric Models* (J. Kmenta and J. Ramsey, eds.) 197–217. Academic, New York.
HILL, B. M. (1988). On the validity of the likelihood principle. In *Statistical Decision Theory and Related Topics IV* (S. S. Gupta and J. O. Berger, eds.) **1** 119–132. Springer, Berlin.
HILL, B. M. (1990). A theory of Bayesian data analysis. In *Bayesian Analysis in Econometrics and Statistics: Essays in Honor of George Barnard* (S. Geisser, J. Hodges, S. J. Press and A. Zellner, eds.) 383–395. North-Holland, Amsterdam.
HILL, B. M. and LANE, D. (1985). Conglomerability and countable additivity. *Sankhyā Ser. A* **47** 366–379.
KOLMOGOROV, A. N. (1950). *Foundations of Probability*. Chelsea, New York.
LINDLEY, D. and SMITH, A. F. M. (1972). Bayes estimates for the linear model. *J. Roy. Statist. Soc., Ser. B* **34** 1–41.
RAMAKRISHNAN, S. and SUDDERTH, W. (1988). A sequence of coin-toss variables for which the strong law fails. *Amer. Math. Monthly* **95** 939–941.
RUSSELL, B. (1937). *The Principles of Mathematics*. Allen and Unwin, London.
SAVAGE, L. J. (1972). *The Foundations of Statistics*, 2nd rev. ed. Dover, New York.
SCHERVISH, M., SEIDENFELD, T. and KADANE, J. (1984). The extent of nonconglomerability. *Z. Wahrsch. Verw. Gebiete* **66** 205–226.
SCOZZAFAVA, R. (1984). A survey of some common misunderstandings concerning the role and meaning of finitely additive probabilities in statistical inference. *Statistica* **44** 21–45.
WHITROW, G. J. (1980). *The Natural Philosophy of Time*, 2nd ed. Oxford Univ. Press.

THE UNIVERSITY OF MICHIGAN
DEPARTMENT OF STATISTICS
ANN ARBOR, MICHIGAN 48109

IAIN M. JOHNSTONE[1]

Stanford University

In Section 5 of his thought-provoking paper, Professor Brown discusses Cox's ancillarity example and draws a distinction between point estimation and confidence procedures. He argues for the conditional validity of his proposed point estimation procedures, since in point estimation no conditionally interpretable stochastic claim is made.

It is, however, possible to make a data-dependent stochastic statement concerning a point estimate without going so far as to provide a confidence set. This may be done by estimating the (squared) error $(\tilde{\alpha} - \alpha)^2$. The issue has been considered in various point estimation settings recently by Rukhin (1988), Lu and Berger (1989) and Johnstone (1988) (the last hereafter denoted by J). Here I shall indicate briefly how some of these ideas extend to Brown's context.

In the setting and notation of Section 3, let $L = L(\{V_i\}, \{Y_i\})$ be an estimate of the squared error $(\delta - \alpha)^2$ of a point estimator $\delta = \delta(\{V_i\}, \{Y_i\})$. The quality of L may be evaluated in turn by using (for simplicity) a quadratic loss $E[L - (\delta - \alpha)^2]^2$, where the expectation is taken over the joint distribution of (V_i, Y_i).

[1]Research supported in part by NSF Grants DMS 84-51750 and 86-00235.