Reply To The Discussion

We are very grateful to the discussants for their stimulating comments. Besides describing interestingly different perspectives, the comments serve to highlight a number of important issues we inadequately discussed in the text.

REPLY TO PROFESSORS BAYARRI AND DEGROOT

It is indeed a pleasure to thank Professors Bayarri and Degroot for their careful reading of our manuscript and the deep insight reflected in their discussion. In the manuscript we tried to explore the implications of the LP and the issues it raises without endorsing any particular mode of inference (until the final chapter); in particular we tried hard not to let our Bayesian point of view color the basic arguments enough to make them unpersuasive to followers of the frequentist tradition. Thus our emphasis was not on "what is the likelihood function?" Rather, we took the likelihood function as given, and argued that the LP would follow no matter what reasonable definition of the likelihood function is used. The definitions in (3.5.1) and (3.5.2) are both reasonable, and serve different purposes.

But we are Bayesians, and are in essentially complete agreement with the basic issues raised by Bayarri and DeGroot. We agree that there is no clear distinction between "parameters" and "variables", and that definition of the likelihood function is ambiguous. As Bayarri and DeGroot observe, any partition of the parameters and variables into two disjoint sets s_1 and s_2, with s_1 containing the observed quantity x, leads to an acceptable likelihood function $\ell_x(s_2) = f(s_1|s_2)$ (providing this function is accepted as "known"). As long as one also keeps track of all known marginal and conditional information about the variables and parameters, any such partition leads to a likelihood function which contains all evidence from the experiment (at least to a

Bayesian). But the need to keep track of this marginal and conditional information, and to treat unknowns in s_1 differently from unknowns in s_2, should be sources of concern to non-Bayesians.

Bayarri and DeGroot suggest that the choice s_1 = "observed" and s_2 = "unobserved" (which they and Professor Kadane call LF_{obs} in BDK) has logical preeminence as a definition of likelihood; then $\ell_x(s_2)$ represents precisely what was learned from the observation of x, unconfounded by any given information about s_2. We again agree; it was only the sociological concerns mentioned in our first paragraph above that kept us from so defining the likelihood function in general.

Further repetition of the insights of Bayarri and DeGroot would be unnecessarily duplicative. Suffice it to say that we agree that non-Bayesians can have a very difficult time defining and interpreting the likelihood function; and once they pass this hurdle, they still must contend with the LP.

REPLY TO PROFESSOR HILL

It would seem rather foolish of us to question Professor Hill's interesting discussion at all, because he seems to feel that we do not go far *enough* in our support of the LP. First we would like to clear up that misimpression (we are fully as enthusiastic as he is concerning the applicability of the LP), and then proceed to the deep issue he raises concerning use of improper, or proper but finitely additive, priors.

From Professor Hill's comments (and also those of Professor Le Cam) it is clear that we did not express ourselves clearly in the Monette-Fraser, Stone, and Stein examples, with regard to the role of frequentist measures and our own conditional perspective of statistical analysis. (This lack of clear expression was primarily due to our concentration on using the examples to indicate the necessity for some type of Bayesian processing of likelihood functions.) Our discussion of frequentist measures was motivated partly by the fact that the examples were historically developed in that fashion, and partly

to indicate that conditional (proper) Bayesians will naturally overcome the
difficulties involved, even from the frequentist perspective. We also do
believe that there can be value in frequentist measures as possible warning
signals that care must be taken in the Bayesian analysis. However, we by no
means meant to imply that (because of a Bayesian-frequentist conflict) one
must *necessarily* change the Bayesian analysis.

These points are, perhaps, best illustrated by the Stein example,
in that it is well recognized that (for data from univariate normal models) the
uniform improper prior is typically quite satisfactory. It is typically satis-
factory, however, *only* because σ is usually small enough that true prior beliefs
can be approximated by the uniform prior. The bad frequentist performance of
the uniform prior in the model (5.3.3) should be a warning that the adequacy of
the uniform approximation to prior beliefs should be investigated, and indeed
such an investigation would usually indicate that the approximation was *bad*;
this would be the conclusion unless y was very small. The frequentist measure,
here, is actually superfluous, however. The conditional Bayesian would
naturally use a uniform prior (as a good approximation) when y/d was very small,
and would recognize (if he had any prior information *whatsoever*) its inadequacy
for typically large y/d, simply because the likelihood function would then be
much more diffuse than even very vague prior information. No knowledge of
frequentist properties, or of differing properties of scale and location
parameters, is necessary to behave sensibly. Also, we in no sense recommend
changing the prior as the model changes. If one's prior opinions truly are
diffuse over the range of the likelihood function, by all means use the uniform
prior, no matter what the model. We simply do not feel that this will be the
case for the model in (5.3.3), however, unless y happens to be exceptionally
small. (Likewise, we judge that the uniform prior will usually be inappropri-
ate for normal models which have monstrously large variances.) There is no
incompatibility with the likelihood principle here, since the "adequacy of the
approximation" can be judged simply by looking at the likelihood function.

The major issue raised by Professor Hill concerns the need for conglomerability, or alternatively the concern that need be felt when the frequentist answer completely contradicts the posterior Bayesian answer. Specifically, in the Monette-Fraser example Professor Hill argues that the uniformly bad frequentist performance of analysis based on the finitely additive uniform prior is not operationally meaningful, because the sample space is, in reality, always bounded. The issue here is not directly related to the likelihood principle, but is another aspect of the possible problems caused by the use of infinite models to approximate reality. If the sample space in the Monette-Fraser example is bounded by N, then certainly the uniform prior becomes permissible, since one can actually simply choose the proper discrete uniform prior on 0 to 2N. We do feel, however, that the subjective assessment of uniformity on 0 to 2N would rarely be reasonable in practice, precisely because the use of the infinite model as an approximation would typically be due to the belief that no X, so large as to be unmeasurable, would actually occur; this *implies* a prior belief that θ could not be extremely large. In general, we would view a uniform conflict between frequentist and Bayesian measures as an indication that either the approximation of an infinite model was inappropriate, *or* the use of the finitely additive prior was inappropriate.

We do, of course, feel that all sample spaces are actually finite, and that (for virtually any problem) one could actually provide a (perhaps overly large) finite sample space. Do examples of the type we are discussing exist for finite sample spaces? If so, such would seem to provide a counter-example to Professor Hill's argument. If not, one could indeed not object, philosophically, to the use of finitely additive measures. There would, however, remain pragmatic questions concerning the practicality of using finitely additive priors (as opposed to countably additive priors) to approximate prior beliefs, but that is an issue for another time and place.

We have long been admirers of Professor Hill's careful treatment of the random effects analysis of variance model, and do not really disagree with his comments here. If we observed a likelihood function, over the range

of which our prior was very diffuse, we would have no qualms about using the
uniform improper prior. If, however, the uniform prior leads to a *procedure*
with bad frequentist properties, we would infer that the uniform prior was a
poor approximation to our prior beliefs for most of the likelihood functions
that could be encountered, and would be loathe to implement it in, say, a
"routine" computer package.

Our view on this matter is partly tied to the discussions surround-
ing Examples 16 and 37. Good frequentist performance will often give some
assurance that a type of conditional Bayesian analysis is moderately robust,
while bad frequentist performance of such an analysis is often an indication of
nonrobustness. Such implications are by no means certain, and use of frequent-
ist verification may often be an inefficient way of investigating robustness,
but we should not dismiss any available aids. In this we also perceive at
least partial agreement with Professor Hill, as witnessed by his numerous
papers on the matter (referenced and discussed in Berger (1984e)).

REPLY TO PROFESSOR LANE

Before considering the two deep issues raised by Professor Lane,
we would raise one minor quibble. His second paragraph consists of a listing
of "bail-out options" for statisticians who choose not to follow the LP. A
major purpose of the monograph was to argue the inadequacy of such bail-outs.
Professor Lane does not make his views on such bail-outs clear, although
presumably, as a Bayesian, he does not accept their validity (for perhaps
reasons other than those given in the text).

The two main issues raised by Professor Lane are (i) the adequacy
of the model paradigm and usefulness of the LP within it, and (ii) the fact
that the LP ignores the Basic Tenet of Bayesianity, namely that inference
should consist of the quantification of uncertainty. In our analyses of these
issues it is particularly important to realize that we perceive little, if
any, disagreement between us and Professor Lane concerning the correctness of
the Bayesian pardigm for statistics. We do differ, however, in our opinions

concerning the most convincing and practically useful way in which this paradigm should be presented. We emphasize the basic agreement because, all too often, non-Bayesians use these rather mild disputes between Bayesians to reject the entire Bayesian paradigm.

Professor Lane only briefly mentions the second issue, that quantification of uncertainty should be the goal of inference. This tenet seems almost self-evident (even though it is not accepted by the bulk of the advanced statistical community), and indeed the LP does not directly incorporate it. An alternate phrasing of the tenet, however, is that statisticians should treat known quantities as fixed and treat unknown quantities probabilistically. The LP does deal with the first half of this phrasing, treating the known data, x, as fixed for inference, while at least treating θ as variable (if not as a random quantity). Hence the LP embodies a major part of the Basic Tenet of Bayesianity.

We have several reasons for approaching the Bayesian paradigm through the LP, rather than through acknowledgment of the Basic Tenet. The first is sociological, and is partly due to the current state of statistics. Two prevelant notions in this "current state" are that the frequentist paradigm provides a satisfactory underpinning for statistics, and that Bayesian analysis is unacceptable because of its prior inputs. It is because of these notions that the majority of statisticians would reject the Basic Tenet, and that direct arguments for Bayesianity often make little headway. Note, however, that the LP directly impunes the first notion, while avoiding the biases of the second notion.

Of course one can argue that transient sociological concerns should not be the basis for judgement, but even from a strictly scientific perspective there is some doubt as to the correct route to take to the Bayesian paradigm. Direct arguments for the Basic Tenet involve some variety of coherency arguments, based on axioms of rational behavior. Such axioms are by no means above criticism. For instance, the arguments listed in Section 3.7, that have been raised against the common "betting scenario," are not easy to dismiss (see

also Le Cam (1977)). Also, even if all such axioms are accepted, the fact
that the only "rational" analyses are those compatible with some Bayesian
analysis does not logically imply that the only acceptable way to do statistics
is to write down a prior distribution (which can never be more than an
approximation to prior beliefs) and perform a Bayesian analysis.

Although the LP is also subject to axiomatic and operational
criticism, it has several advantages in these regards. The first is that its
axiomatic basis is compelling to most people. The WCP is compelling to almost
everyone, and the SP is an integral part of most existing statistical paradigms.
We do acknowledge that the SP is not really "obvious," and indeed went to
considerable effort in Sections 3.6 and 3.7 to justify the principle (and not
just for the "betting scenario"). The simple fact remains, however, that very
few statisticians will reject either of these axioms, while most seem unmoved
by the coherency axioms.

As to the operational criticism, the LP would again seem to have
an edge, precisely because it does *not* provide a final answer and can hence be
more specific in its *partial* answer. The coherency approach provides only the
vague general requirement that substantial inconsistency with some Bayesian
analysis should be avoided. The LP is, on the other hand, specific in its
recommendation to utilize only the observed likelihood function, even though it
does not address the question of how this is to be done. And from a purely
pragmatic viewpoint, this first step may well be the most important step of
all. The reason is that, in practice, one often spends the greatest effort in
model selection and verification; the knowledge that one need only consider
the *observed* likelihood function can simplify this task enormously. Indeed,
it is not unusual for the choice of a prior on model parameters to be of such
secondary importance that one never gets beyond "playing with likelihoods."

Professor Lane does point out that the "standard" decomposition in-
to model and prior is often artificial, and so should not be a part of statis-
tical foundations. While sympathetic, we view such decompositions as

essential practical simplifying devices, necessary to achieve progress. One
usually progresses on a hard problem by identifying simple components that can
first be analyzed separately, and then combined together. Although we agree
that such decompositions are not always appropriate, their pragmatically
central role to statistics is hard to deny. And the fact that the LP provides
so much insight into what is probably the most crucial component of the decom-
position, gives it considerable appeal.

In summary on this issue, the coherency approach to the Bayesian
paradigm has many admirers (ourselves among them) and can perhaps claim a
logical ascendancy over the LP approach, but (for the reasons mentioned above)
we feel that the LP approach has had, or at least can have, a larger impact.
The quotation on p. 2 from L. J. Savage is revealing in this regard, coming
from an ardent admirer of coherency. (More complete discussions of this
issue can be found in Berger (1984b) and Berger (1984e).)

It was perhaps unfair to spend so much time on this issue, given
that Professor Lane only briefly mentions coherency. However, we feel that it
is important to view Professor Lane's objections to the LP in the larger per-
spective of alternative approaches to the Bayesian paradigm.

Let us now turn to Professor Lane's second issue, specifically
the criticisms about the "model" paradigm and the applicability of the LP
within it. The first issue Professor Lane raises is that of the interpretation
of θ, and the question of applicability of the LP unless θ is a "real"
physical parameter. We wanted to avoid the philosophical problems inherent in
any discussion of the meaning of parameters, but in retrospect should have
spent more time on the issue. The reason is that, while of course the LP will
apply if θ is a real physical parameter (in some sense), it also applies in
the much more common situation where θ is only defined by some aspect of the
experiment. For instance, a very large part of statistics deals with situa-
tions involving a series of (approximately) i.i.d. observations X_i. The
parameter θ is often implicitly defined by the assumed density (say), $f_\theta(x_i)$,
for the observations, and is *not*, as Professor Lane implies, necessarily

identified with the overall $\{P_\theta\}$, which could also involve other aspects of the experiment such as the stopping rule, possible censoring, and so forth. We could consider any number of experiments with the same implicitely defined θ, but with different $\{P_\theta\}$, and apply the WCP and SP to deduce the LP. Indeed a major purpose of the LP is to show that features of $\{P_\theta\}$ which are *irrelevant* to the implicit definition of θ are ignorable in the analysis. Professor Lane's three interpretations of θ do not cover this case, which we would call the case of major practical interest.

Of course, we did not mean the LP to apply in Professor Lane's case b), where Θ is just an index set, and specifically warned against this on several occasions. (The entire mixing setup makes no sense if the parameters in the two experiments can differ.) Our failure to carefully define θ in examples was admittedly sloppy, but was based on the desire to avoid complex philosophical issues that are of uncertain practical import. (Convincing a practicing statistician, who routinely uses models, to base his analysis on the observed likelihood function is a significant practical step. Informing him that his model parameters really have no meaning is unlikely to cause much improvement in his statistical practice.)

Professor Lane next questions the value of inference about model parameters, arguing that predictive inference about future observables is of most concern. We do not dispute this point, but, as Professor Lane acknowledges, we do handle predictive inference by incorporating the future observable in θ. The complaint that the LP does not then say how to eliminate θ is one of the arguments we use for Bayesian implementation of the LP, but the complaint in no way limits or casts doubt on the LP. Professor Lane may prefer the de Finetti approach, which allows direct dealing with predictive inference, but, as discussed earlier, we feel the model-based "half-way house" is generally a pragmatic necessity. It is enormously difficult to attempt directly to ascertain such complicated things as predictive distributions. Even inventing crude models and artificially creating model-prior separations will, we feel,

serve predictive statisticians best in the long run. Professor Lane does raise the valid point that our emphasis throughout the text on model parameter inference, itself, may be misleading. Our only defense is the essential impossibility of sensibly discussing predictive inference outside a Bayesian framework, combined with our desire to minimize Bayesian involvement (for already mentioned sociological reasons).

We have tried to describe accurately the reasons for our preference for the LP approach to Bayesianity. Admittedly this preference may be due to our traditional probabilistic and statistical background (with its model orientation), but, on the other hand, the alternative developments have not managed to produce any broadly useful new practical methodology. There is real danger in letting philosophical games obscure the *practical* realities of the situation. (For instance, the coherency game of "betting" serves to give various sound meanings to probabilities, but it seems completely backwards in its application: people decide how to bet by *first* determining probabilities, usually through some comparative likelihood method.) A philosophical game that can be played to support the LP, foundationally, is the "finite sample space" game (see Section 3.6.1). Reality always has a finite sample space, and the LP always applies to the implied "model." This formulation has little operational significance, however, and so we do not view it as a serious argument for the LP approach.

In conclusion, although we certainly support, and indeed find philosophically enlightening, approaches to Bayesianity based on coherency, our own preference is for the LP approach.

REPLY TO PROFESSOR LE CAM

The major and probably most important point made in Professor LeCam's interesting discussion is that we should be "a bit unprincipled." He sees value in both classical methods and the LP. As "tools" in the statistician's toolkit, we agree that there is possible value in classical methods, although we would tend to prefer Bayesian tools, if available. The choice of

a tool is, however, not really the question addressed in the monograph. The
purpose of the monograph was to attempt to clarify the more fundamental
question: What should the statistician be using his tools for? We believe
the vast majority of statistical users want to know "the evidence about θ from
E and x," and indeed will likely be unable to assign any other meaning to a
statistical conclusion. Because of the demonstrated conflict between this
goal and the frequentist goal of procedure performance (except, of course, for
the various discussed exceptions, such as experimental design), we feel that
this question of *purpose* can not be ignored. And while a variety of tools
may be useful in reaching our stated goal of the determination of conditional
evidence (even frequentist tools may be useful - see Section 5.4 and also
Berger (1984b) and Berger (1984e)), we would argue that the value of the tool
must be related to this ultimate goal. The big stumbling block in the long-
running controversy in statistics has been the lack of separation of purpose
and method.

        A recurring theme in Professor LeCam's discussion is the issue of
communication of statistical evidence. Indeed, because we briefly indicate
in Chapter 5 that we feel it necessary to be Bayesians (and hence produce
priors and posteriors), Professor LeCam intimates that we have "argued...into
a corner." Our interpretation, however, is that, even if communication of
evidence through Bayesian measures is deemed unappealing, it is a *scientific
necessity*, unless one is willing to sacrifice the goal of communicating the
actual evidence obtained about θ. Of course, the Bayesian situation (as
regards scientific communication) is not nearly as bad as many non-Bayesians
think; the spectre of being forced to accept someone's unreasonable prior
distribution is not really an issue. Good Bayesian reporting can be done
with a variety of strategems involving the presentation of the conclusion for
a wide variety of priors (c.f. Dickey (1973)). And simply presenting likeli-
hood functions or, perhaps somewhat better, posterior distributions for
noninformative priors can be viewed as a reasonable conditional communication
device. Of course, such are not traditional in scientific journals, but

we all know of a number of "traditions" concerning statistical reporting in
scientific journals that we would all gladly retire.

As to Professor LeCam's feeling that one should report all possibly
relevant data about an experiment, no subjectivist would think of disagreeing.
After all, a subjectivist is (theoretically) responsible for producing his
likelihoods (as well as priors), and *all* data about the experiment could be
relevant to this enterprise. Of course, the LP does say that in processing
all this information the conditional viewpoint should have primacy.

Professor LeCam feels that a major flaw in the LP axiomatics is the
assumption that Ev(E,x) exists. Since we allowed Ev(E,x) to be *anything*, any
collection of conclusions or reports, we are unclear as to the exact objection.
(One surely must make some report.) All the axiomatics say is that if one
processes information in violation of the LP, perhaps by reporting frequentist
error probabilities, then one is behaving in violation of either the WCP or SP
or both. It is, perhaps, conceivable that, for each experiment, one could
process information in a completely new way, so that one's Ev(E,x) would be
continually changing, and so that no violation of the WCP or SP could be
established. This, however, is not realistic: as statisticians we are bound
to standardize many of our analyses, or at least parts of many of our analyses.
The text argues that any such standardized methods of processing information
should be in accord with the LP.

Professor LeCam is certainly correct in his comment that our
passage from the LP to Bayesianity is much weaker than the argument for the
LP. We felt little need to rigorously justify this final step, mainly because
we feel that it is belief in the LP that is the major hurdle; it is hard to
avoid becoming a Bayesian after fully accepting the LP.

Professor LeCam feels that we make a direct appeal to frequentist
ideas in our attempt to resolve the Stein example. We clearly did a bad job
in the example, of explaining our position, because Professor Hill likewise
sees us as resorting to frequentist reasoning. The passage to which
Professor LeCam refers was an attempt to explain *to frequentists* why, as

Bayesian conditionalists, we would not suffer from a frequentist perspective. The conditional Bayesian analysis we discussed in no way depended on frequentist evaluations, however. For a more lengthy discussion of this point, along with a brief description of the role conditional Bayesians can ascribe to frequentist measures, see our reply to Professor Hill.

The final issue raised by Professor LeCam is that of application of the LP when only "approximate likelihoods" are available. We have seen no evidence to indicate that the need to use approximations with the conditional approach causes any more problems than the use of approximations with any other approach. In the example of n independent Cauchy observations, we would of course prefer use of the exact observed likelihood function, but if n were enormous and we had technical problems in calculating and using the exact likelihood, we would certainly consider using the $\eta(\theta, \frac{2}{n})$ approximation. But we would use the *observed* likelihood function from this approximation as the experimental input to evidence, not frequentist measures calculated by averages over the normal approximation. Without knowing the specific problem one cannot safely recommend specific priors. When n is large, however, prior information will typically be vague compared with the likelihood function, so use of the noninformative uniform prior would be a reasonable first approximation.

REPLY TO PROFESSOR LE CAM'S SECOND EDITION COMMENTS

We are sympathetic to Professor LeCam's position, that attempting to summarize a complex situation by the pair (E, x) may omit much that is relevant. Thus we have always been interested in attempts to depart from the usual statistical framework of probabilistically-modelled experiments (though we have yet to see an alternative framework that works better). Note, however, that virtually all of classical statistics is based on considering particular notions of Ev (E, x). Thus Professor LeCam's observations would seem to apply equally well to all standard statistical concepts. He does mention the possible need for "introducing in the system a variety of concepts that go beyond pairs (E, x)"; the argument to abandon (or at least extend) the usual statistical

framework is too big for us.

Perhaps Professor LeCam is making the smaller logical point that principles (e.g., the LP) that are deduced within a too narrow framework are not necessarily valid in the correct framework. The constructive side of the LP ($\ell_x(\theta)$ summarizes what is needed from (E, x)) may thus be questioned; but the destructive side of the LP, that measures based on (E, x) which are incompatible with the LP (such as frequentist measures) are contraindicated, seems intact. After all, a frequentist measure based on (E, x) should certainly be able to pass an evaluation in its own domain. If it fails there, it is hard to imagine that it would be good in an enlarged domain.

We have been a bit overly dogmatic to emphasize our basic views. At the same time, our position, stated in Sections 5.4 and 5.5, bears a certain similarity to LeCam's, in that we also do not feel that all our actions "*must abide by the LP.*" Our own summary position, however, is that abiding by the LP is a generally good guideline, and that major deviations from the LP are highly suspect.