Why Bayesianism?[1]
A Primer on a Probabilistic Philosophy of Science

Several attempts have been made both in the present and past to impose some *a priori* desiderata on statistical/inductive inference (Fitleson, 1999, Jeffreys, 1961, Zellner, 1996, Jaynes, 2003, Lele, 2004). Bringing this literature on desiderata to the fore, I argue that these attempts to understand inference could be controversial. This is why Royall's (1997, 2004) views on the foundations of statistics are more fruitful. Royall distinguished among three types of questions, (i) the *belief question,* (ii) the *evidence question* and finally (iii) the *acceptance question* (van Fraassen 1991). He thought that Bayesians could only handle the first question, whereas classical statistics (error-statistics), can address the third question. He contended why the Likelihood framework alone is able to answer the second question.

Royall's work makes it clear that statistical inference has multiple goals. As a result, it is unlikely that one measure is able to address all problems in statistical inference at the core of a probabilistic philosophy of science. The purpose of the paper is to evaluate Royall's work from a Bayesian perspective. Contra him, I contend that Bayesianism and Bayesianism alone is able to address all three questions in a manner that is at least as satisfactory as classical statistics (error-statistics) or likelihood approach. So the answer to the question, "Why Bayesianism?" is that Bayesian School alone provides a unified approach to probabilistic philosophy of science.

Prasanta S. Bandyopadhyay
Department of History & Philosophy
Montana State University, Bozeman,
MT 59717, USA
Email: psb@montana.edu

[1] The author wishes to thank Jose M. Bernardo, Robert J. Boik, Gordon Brittan Jr., Colin Howson, Jayanta Ghosh, Mark Greenwood, Subhash R. Lele, Sue Monahan, James Robison-Cox, Tasneem Sattar, Elliott Sober, C. Andy Tsao and an anonymous referee for suggesting several improvements in the contents of the paper by their direct or indirect comments regarding the issues raised here. The author is especially indebted to both John G. Bennett and Mark L. Taper for numerous discussions/comments regarding the various aspects of the paper. He also thanks the participants of the Philosophy of Science Association meetings in the workshop on Royall's work held in Milwaukee 2002 and the International Workshop/Conference on Bayesian Statistics held in Varanasi 2005 for their comments and encouragement.

Overview

The answer to the question, "Why Bayesianism?" is that the Bayesian School alone provides a unified approach to the statistical inference at the core of scientific methodology. Deductive and inductive arguments are essential aspects of scientific methodology. An argument is deductively valid if and only if it is logically impossible for its conclusion to be false whenever its premises are true. A deductively valid argument is distinguished from its inductive counter-part by the property of "monotonicity." The monotonicity property states that one can't undermine the conclusion of a deductively valid argument by adding a new premise provided all the original premises remain intact. Inductive inference does not satisfy the monotonicity property (Adams, 1998; and Bandyopadhyay and Bennett, 2004). Nonetheless, it is inductive inference that underlies much of statistical inference.

Unlike the conclusion of a deductive argument, the conclusion of an inductive argument does not follow necessarily from its premises. As a result, several attempts have been made both in the present and past to understand and ultimately, if possible, to develop a "logic" analogous to deductive logic for inductive arguments. Rudolf Carnap (1952), for example, searched for general inductive rules by which, given information about past results, one could make the best possible prediction of future events. In spite of Carnap's impressive work, his program was universally judged as failing to deliver such rules. Although statisticians and philosophers have abandoned the Carnapian research program, many still believe in the need for imposing some *a priori* desiderata on statistical/inductive inference (Jeffreys, 1961; Zellner, 1996; Zellner, 1997; Jaynes, 2003; Lele, 2004; Christensen, 1999; and Fitleson, 1999). I argue that these attempts to understand inductive inference fail to do justice to various dimensions of inductive inference. The justification for satisfying specific desiderata is contingent on which goals have been adopted, and desiderata could vary from one agent to another. This is why Richard Royall's (1997, 2004) views on the foundations of statistics, which take account of this fact, are more fruitful.

Royall distinguished three types of questions. He calls the first two, (i) the *belief question*, and (ii) the *evidence question*. Lastly (iii), borrowing an expression from philosophical literature, I will call Royall's third question the *acceptance question* (van

Fraassen 1991). Royall thought that Bayesians could only handle the first question, called the belief question, whereas classical statistics (error-statistics) can address the acceptance question. Royall contended that the likelihood framework alone is able to answer the evidence question. The notion of likelihood is used to answer the question, "how likely are the data given the hypothesis?" Those who believe that the notion of likelihood alone is adequate to capture statistical inference belong to the likelihood framework.

Royall's work makes it clear that statistical inference has multiple goals. As a result, it would seem unlikely that one account is able to address all problems in statistical inference. In response to Royall, I have developed three accounts to satisfy these multiple goals within Bayesianism. They are (i) the confirmation account, (ii) the evidence account, and (iii) the acceptance account. Contrary to Royall, I contend that Bayesianism and Bayesianism alone is able to address all three questions in a manner that is at least as satisfactory as classical statistics (error-statistics) or the likelihood approach.

The central theme of this paper is to bring Royall's three questions to a larger audience and then provide a Bayesian evaluation of these questions. In light of this theme, the paper is divided into nine sections. In the first section, I will broach Royall's three questions. In the second section, I will discuss why Bayesians could be taken to address solely the belief question by drawing on the works of Harold Jeffreys and Arnold Zellner, two kindred-spirited Bayesians who hold the same kind of views on inductive/statistical inference. In the third, I will discuss Subhash Lele's recent non-Bayesian approach to statistical evidence and point out that his work is exclusively tied to the evidence question. In the fourth section, I will address why classical statistics deals only with the acceptance question. The next section will be devoted to evaluating both Bayesian and non-Bayesian approaches to inductive/statistical inference. Here, I will argue that Royall's emphasis on multiple goals is crucial in understanding foundational questions in statistics. In the sixth section, I will propose two accounts, the confirmation account and the evidence account as part of a Bayesian response to Royall's belief and evidence questions. Furthermore, I will detail some crucial unnoticed differences between an account of confirmation and an account of evidence. In the seventh, I will revisit the belief and evidence questions with diagnostic examples involving simple hypotheses. Here, I will also discuss, among other things, Deborah Mayo's error-statistical account (Mayo, 2004). The eighth section will address one model selection problem involving cases that are not liable to the same treatments that simple hypotheses will receive in section seven. I will propose a criterion called the Bayes' Theorem Criterion (BTC) to address the problem in model selection. In section nine, I will develop a decision theoretic account of acceptance. Finally putting the pieces together, I will argue the merit of having a unified account and why Bayesianism alone is able to provide that unification adequate for a probabilistic philosophy of science.

1

Royall's Three Questions

Royall argues that one needs to distinguish three types of questions pertaining to three problem areas in statistics. Consider two hypotheses, H, representing that the patient suffers from cervical cancer, and ~ H, its denial. Assume that a PAP smear test, which is administered as a routine test for screening for cancer, comes out positive for the patient. Given the datum that the test comes out positive, one could ask three questions:

(i) Given the datum, what should we *believe* and to what degree?
Or,
Given the datum, how should we *change* our degree of belief?
(ii) What does the datum say regarding *evidence* for H against its counter-parts?
(iii) What should we *do*?

The first question Royall calls the *belief question*, the second one the *evidence question* and the third one is what I call the *acceptance question.* He further argues that, depending on which school of statistics one belongs to, one asks the particular question appropriate for that school. He thinks that Bayesians ask the belief question, likelihoodists like him ask the evidence question, and classical statisticians ask the acceptance question. Within the likelihood framework, Royall has done three things. (1) He has clearly spelled out what is really the evidence question. (2) He argues that the likelihood ratio (LR) is the correct measure of evidence. Finally (3), he contends that the law of likelihood (LL) justifies the use of likelihood ratio as a measure of evidence. According to the LR, D is evidence for H over ~H2 if the ratio Prob(D|H)/Prob(D|~H) is

> 1, whereas the LL says that if Prob(D|H) > Prob(D|~H) then D is evidence for H over ~H.

2

The Belief Question

I will address Jeffreys and Zellner's Bayesian approach to imposing desiderata on a statistical/inductive theory and conclude that they are concerned with the belief question.

Although Jeffreys wrote before Zellner, they share a great many common convictions about how and why a Bayesian theory needs to be developed. Both believe that "Bayesian statistics is the technology of inductive inference." They know the fundamental difference between inductive and deductive inferences. They share similar views about imposing *a priori* constraints on inductive/statistical inference to make it more effective, especially in scientific practice. For the sake of brevity, I address those aspects in which their views are similar. I call their views the Jeffreys-Zellner approach (JZA) to inductive/statistical inference.

The JZA consists of eight *a priori* desiderata. The first five Jeffreys consider "essential" and the rest of them he thinks are "useful guides." Here are Jeffreys-Zellner's eight rules.

(i) All hypotheses should be clearly stated and the conclusions should follow deductively from these hypotheses.
(ii) A theory of induction should be self-consistent implying that it is logically impossible to derive contradictory claims from that theory along with a given set of observations.
(iii) Any rule should be applicable in practice. Any definition becomes useless unless the object of a definition can be recognized in terms of the definition when it occurs. Any reference to an existing thing or the estimate of a quantity should involve an empirically possible experiment.
(iv) A theory of induction should make room for the possibility that a particular inference using the theory might turn out to be incorrect.
(v) A theory of induction should not eliminate any empirical proposition *a priori*.
(vi) The number of postulates in a theory should be as few as possible.
(vii) Although a theory of induction need not regard a human mind as a perfect reasoner, it should at least agree with actual thought process in outline.
(viii) Because of the vast complexity of inductive inferences, one should not hope to develop a theory of inductive inference as thoroughly as a theory of deductive inferences.

Consider now how these *a priori* constraints help both Jeffreys and Zellner to develop a theory of inductive inference, especially of a Bayesian kind. Zellner concurs with Jeffreys when the latter writes,

> They [these eight rules] rule… out any definition of probability that attempts to define probability in terms of infinite sets of possible observations, for we cannot in practice make an infinite number of observations. The Venn limit, the hypothetical infinite population of Fisher and the ensemble of Willard Gibbs are useless to us by [rule iii].

To explain why Jeffreys thinks that the rule eliminates the Venn limit, I will discuss how he looks at it. According to the Venn limit, if the number of trials tends to infinity and an event in these trials occurs a large number of times, then the probability p is the limit of the ratio of the number of times when p will be true to the whole number of trials. Jeffreys thinks that the defect of this definition is that it is inapplicable in nature because the presupposition of the limit eliminates total randomness in nature. In addition, he argues that the Venn limit goes against the rule (iv) because the former restricts *a priori* how nature should behave.

He goes on to explain why he thinks that the Venn limit is inapplicable in scientific practice. He writes,

> "The form of this definition restricts the field of probability very seriously. It makes it impossible to speak of probability unless there is a possibility of an infinite series of trials. When we speak of the probability that the Solar System was formed by the close approach of two stars, or that the stellar Universe is symmetrical, the idea of even one repetition is out of question; but it is just in these cases that the epistemological problem is most acute. But this is not all, for the definition has no practical application whatever." (Jeffreys, 1957; p 182).

In contrast, both Jeffreys and Zellner attribute probability statements to a parameter. Zellner writes, "The operations of Bayesian statistics enable us to make probability statements about parameters' values and future values of variables." (Zellner, 1984; p.6). Jeffreys thinks that the Venn limit, the hypothetical infinite population of Fisher and the like, rest on a notion of probability which is frequentist in spirit. Spelling out rule (iv), he argues that our state of knowledge about scientific laws, including the relativity and quantum theories, provides strong reason to believe that there is no conclusive evidence to consider that our present laws are final. Since we may not able to give a deductive proof or disproof for those theories, Jeffrey writes, "there is a valid primitive idea of expressing the degree of confidence [degree of belief] that we may have in a proposition." He goes on to state that "*the essence of the present theory* is that no probability… is simply frequency. The fundamental idea is that of a reasonable degree of belief, which satisfies certain rules of consistency and can in consequence of these rules be formally expressed by numbers….."

For Jeffreys, then, the notion of probability is an agent's degree of belief in a proposition/hypothesis where the agent's degree of belief must be coherent. For them, this coherence condition implies that the agent's degree of belief must satisfy the rules of the probability calculus along with the rule 6, which is known as the simplicity postulate. According to this postulate, "the simpler theories have the greater prior probabilities." Zeller has recently extended this simplicity postulate to various models of economics (Zellner, 2001). Given the JZA, the only plausible theory turns out to be the Bayesian theory of inference that makes central the agent's degree of belief in a parameter or a hypothesis satisfying some coherence conditions. So Royall is right in pointing out that the way Bayesianism has been practiced, given the JZA, one could only obtain the agent's degree of belief central to Bayesianism.

However, one key feature of Bayesianism needs to be mentioned.[2] A crucial aspect of statistics, possibly significantly more crucial than parameter estimation, is prediction. Bayesians are in a much better shape here than frequentists, because the former can incorporate the uncertainty due to estimation of parameters. Frequentists use only plug-in predictors which can't address the said problem.

3
The Evidence Question

Now consider Lele's non-Bayesian approach to desiderata for a statistical theory of evidence. He is a non-Bayesian because he works within an evidentialist framework (that is, a generalized version of the likelihood framework) that eschews use of subjective prior probability. Lele begins with the law of likelihood and then defines a class of functions called "the evidence functions" to quantify the strength of evidence for one hypothesis over the other. He imposes some desiderata on this class of evidence

[2] I owe this point to an anonymous referee.

functions, which could be regarded as epistemological conditions. Some of the conditions satisfied by the evidence function are

(i) The translation invariance condition: If one translates an evidence function by adding a constant to it to change the strength of the evidence, then the evidence function should remain unaffected by that addition of a constant.

(ii) The scale invariance condition: If one multiplies an evidence function by a constant to change the strength of evidence, then the evidence function should remain unaffected by that constant multiplier.

(iii) The reparameterization invariance: The evidence function must be invariant under reparameterization. It means that if there is an evidence function, Ev1 and the latter is reparameterized to Ev2, then both Ev1 and Ev2 must provide the identical quantification of the strength of evidence.

(iv) The invariance under transformation of the data: The evidence function should remain unaffected insofar as the quantification of the strength is concerned if one uses different measuring units.

Lele has also added two more conditions on the evidence function, which he calls "regularity conditions," so that the probability of strong evidence for the true hypothesis should converge to 1 as the sample size increases. He shows that the likelihood ratio becomes an optimal measure of evidence under those epistemological and regularity conditions, providing a justification for the use of likelihood ratio as a measure of evidence. He believes that showing the optimality of the likelihood ratio amounts to providing necessary and sufficient conditions for the optimality of the law of likelihood.

One could find a flaw in Lele's approach, because the law of likelihood is faulty. Consider an example. A fair die has been thrown, but we do not know which face is up. The investigator has proposed H1 and H2 separately as two likely hypotheses to explain the true nature of the die. H1 = 1 is up; H2= 6 is not up; and the data (D) is the number of up which is not greater than 2. Based on the law of likelihood, the agent is forced to conclude that D provides more support for H1 over H2 because the probability of the data given H1 is 1, whereas that of the latter is 2/5. However, if D is true then it follows deductively that H2 must be true. The relationship between D and H2 is a logical relation, therefore, Prob(D|H2) must be 1. No legitimate inference could be made from H1 to H2 insofar as their comparative evidential support relative to the data is concerned (See section VI for more.) The law fails to respect this deductive relationship between D and H2 yielding an erroneous conclusion (See Levine and Schervish, 1999; Sober, 2005. However, my example is due to John Bennett.) This counter-example to the law raises a problem for Lele's argument that the likelihood ratio is an optimal evidence function, because, in the evidentialist framework, the law justifies the use of likelihood ratio in the first place.

One could, however, appreciate the theme of the section while overlooking the defect of the law of likelihood. Consider how Lele has chosen regularity conditions that would yield the likelihood ratio as an optimal evidence function. Lele is aware of this and writes, "Toward this goal, some additional regularity conditions are imposed." It is interesting to note that those conditions on a class of evidence functions are quite general because they could yield the Kullback-Leibler disparity measure (see Burnham and Anderson, 1998) as an instance of that evidence function. However, they eliminate the posterior distribution as an instance of this evidence function. In fact, the parameter transformation invariance is not satisfied by the posterior distribution. That is, the latter violates condition (iii) on this class of evidence functions.

In our modeling of nature, the same phenomena may be represented equally well by different formalisms. It is arbitrary whether we quantify spread by the variance, v, or by the standard deviation, s. Exactly the same information is conveyed by these two parameterizations because there is an exact transformation between them. Bayesian inference, however, is not invariant under parameter transformation. At the core of Bayesian inference lies posterior probability distribution. A Bayesian takes an estimate to be of some function of a mean or mode of a posterior distribution. For discussion, let us say that the researcher chooses to use the mean of the posterior distribution of a parameter as a point estimate. One can transform a probability distribution, f, for the parameter v into a probability distribution, h, for the parameter s using the idea of change of variable technique (e.g. Casella and Berger, 1990). The mean of f(v) need not be equal to the square of the mean of h(s) within a Bayesian framework. As a result, the Bayesian point estimates for the two parameterizations need not match.

Consider an example to see why the Bayesian point estimates for the two parameters may not yield the same conclusion after transforming one parameter into another via change of variable technique. Suppose a set of data is distributed with log-normal (m, sd) in which sd = 2 and m = 2. Based on this information about the distribution of the data, we would like to make an inference about the parameter v. As a Bayesian, we need to have a posterior distribution of v, which is as follows:

$$f(v) = \frac{\exp\left[\frac{-(\ln(v)-m)^2}{2sd^2}\right]}{vsd\sqrt{2\pi}} \quad .$$

Because the above function along with other functions applied in this specific example are not defined at zero, I have used numbers different from zero as their lower limits in all my integrations.

The value of the integration, $\int_{.000001}^{\infty} f(v)dv$ =1 proves that it is a probability distribution. Now by using the same distribution of the data, we would like to make an inference about the parameter s, which is a posterior distribution of s derived via change of variables.

$$h(s) = \frac{|2s|}{s^2\sqrt{2\pi sd^2}} \exp\left[-\left[\frac{(\ln(s^2)-m)^2}{2sd^2}\right]\right]$$

$\int_{.001}^{\infty} h(s)ds = 1$. This turns out to be a probability distribution too. The mean of the posterior distribution for s is

$$mnh = \int_{.00001}^{\infty} sh(s)ds \qquad \text{mnh}= 4.482$$

In contrast, the mean of the posterior distribution for v is

$$mnv = \exp\left(m + \frac{sd^2}{2}\right); \int_{.000000001}^{\infty} vf(v)dv = 54.598 \rightarrow \qquad \text{mnv} = 54.598 \rightarrow \sqrt{mnv} = 7.389.$$

These two values, 4.482, which is the mean of the posterior distribution for the parameter “s” via the probability distribution h, and 7.389, which is the mean of the posterior distribution for the parameter “v” via the probability distribution f, are not equal.

What this example shows is that if one imposes a desideratum on the evidence function that the latter must satisfy the reparameterization invariance property, then one won’t be able to use posterior distribution, a pillar of Bayesianism, as a suitable evidence function.[3] I will argue (section VI) that posterior distribution provides a framework for answering the belief question, which is different from the evidence question. So Royall is right that the likelihood framework or the evidentialist framework, which is a generalized version of the likelihood framework, is able to answer the evidence question, and not the belief question.

4
The Acceptance Question

According to Royall, classical (error-statistics) is decision-theoretic in character; as a result it is interested in the acceptance question. Classical statistics has to do with computing error probabilities that are either tied to the data and particular experimental set-ups or generated by human intervention in these set-ups. Classical Neyman-Pearson statistics neatly summarizes these two types of errors as Type I and Type II. Suppose the clinician is interested in knowing the effect of the drug AZT among patients suffering from HIV after it has been administered (Neyman, 1967.) To learn the effect of the drug, and to understand how the situation is related to the two effects of errors, consider two hypotheses: Ha and H0. Ho is the null hypothesis; it is that there is no difference in patients before and after treatment. Ha is the alternative hypothesis that there is a difference. The Type I error occurs when the decision is to reject the null hypothesis when it is actually true. The Type II error occurs when the decision is not to reject the null hypothesis when it is actually false. The *goal* of error-statistics is thus twofold: to reject the null hypothesis as a function of some pre-assigned significance and to minimize

[3] One might worry how effective this objection to Bayesianism is when one argues that posterior distribution does not remain invariant under parameterization. The reason for this worry might be that even within a frequentist paradigm, some properties of an estimate are also *not* invariant under transformation. Even though T is an unbiased estimator of θ, T^2 might not be an unbiased estimator of θ^2. If within a frequentist's framework an unbiased estimator of θ could become a biased estimator of θ^2 after transformation, then, the objector wonders, why does the failure of posterior distribution to satisfy parameter invariance should be touted as a defect for Bayesians? One response is that although the argument that the failure for posterior distribution to satisfy parameter invariance is related to the above point regarding the failure to satisfy unbiasedness of a parameter after transformation, the former is a different objection than the latter one. For example, although MLE is not necessarily an unbiased estimator, the information contained in MLE point estimate remains unchanged under reparameterization. To put the latter point formally, MLE of V = (MLE $\sigma)^2$. The point is that whether two estimates yield equivalent value under parameterization is different from whether that estimate is an unbiased estimator. I thank both an anonymous referee and Mark Taper for clarifying two related issues here.

these two errors. Now whether Mayo would accept this characterization of error-statistics as decision-theoretic is another story. However, it seems clear that Royall is right in pointing out this decision-theoretical character of the notion of acceptance in the case classical statistics.

From this discussion, one pattern of reasoning emerges insofar as (especially) both the Bayesian and likelihood/evidentialist approaches to desiderata are considered. Whether one is a Bayesian or non-Bayesian, one would like to impose those *a priori* constraints on one's methodology that would yield only one's favorite theory, Bayesian or otherwise. I have only provided two cases in which this pattern is realized. However, this pattern is typical among both philosophers and some working statisticians.

5

Three Questions, Different Desiderata, and Various goals

There could be several desiderata for a model/theory to be good. However, working scientists are well aware that one might not be able to capture all desiderata of a model when confronted with a hard problem. The curve-fitting problem, which is an instance of the model selection problem, provides an example of this kind. In the curve-fitting problem, the investigator is confronted with how to fit a line to a set of data. If she would like to capture the data well, then she would have a high goodness-of fit value, which is one desideratum for a model, but the mathematics of that line could get complicated. In contrast, if she would like to make the model simple, then she would prefer a straight line, thus failing to capture the data well. So the curve-fitting problem arises because one can't simultaneously optimize disparate desiderata such as "goodness-of-fit" and "simplicity." Even knowing this conundrum of model selection well, why do working scientists, Bayesian and non-Bayesian, search for some a priori constraints on their theories? The rationale is that they have different goals, which they would like to achieve with the help of those desiderata that would not generate a tension in theory selection. To escape this possible tension stemming from conflicting desiderata, Jeffreys and Zellner have adopted (ii), that is, a theory of induction should be self-consistent where it is logically impossible to derive conflicting results from that theory along with a given set of data.

This Bayesian/non-Bayesian attitude to desiderata is not new in scientific methodology. It dates back to Descartes who recommends rules for how one should conduct a correct discourse in epistemology. Descartes in his *Discourse on the Method* (VI 18-19, HR 1 92) lays down four rules that should precede any methodology.

"(i) Accept nothing as true which I didn't clearly recognize to be so: that is to say, to avoid carefully precipitation and prejudices, and to accept nothing in my judgment beyond what presented itself no clearly and distinctly to my mind, that I should have no occasion to doubt it.
(ii) Divide each of the difficulties which I examined into as many parts as possible, and as might be necessary in order best to resolve them.
(iii) Carry on my reflections in order, starting with those objects that were most simple and easy to understand, so as to rise little by little, by degrees, to the knowledge of the most complex: assuming an order among those that did not naturally fall into a series.
(iv) Last, in all cases make enumerations so complete and reviews so general that I should be sure of leaving nothing out."

It is no wonder that Descartes himself has followed these rules in building a foundation for his epistemology. Commenting on Descartes' rules of method, Leibniz gibes that they were 'like the precepts of some chemist; take what you need and do what you should, and you will get what you want.' (*Philosophical Schriften von G.W. Leibniz*, ed. C. I. Gerhardt, Berlin, 1875-90, vol. IV, p. 328). My own position on imposing *a priori* constraints is like the position Leibniz has adopted toward evaluating Descartes. Leibniz criticized Descartes for recommending that methodologists follow those rules. He argued that since Descartes had a goal to build a foundation of knowledge, Descartes needed those rules badly. However, if somebody has a different goal, she might ask for different rules to follow. Leibniz concludes that satisfying those desiderata are solely contingent on what one would like to achieve. As a result, those desiderata could vary from one agent to another. This variation in desiderata is likely to be due to some specific questions/problems investigators are confronted with. This is why I believe that the way Royall has addressed the foundations of statistics is much more fruitful.

The significance of these questions is that, unlike Jeffreys, Zellner and Lele, Royall makes it explicit that statistical inference has multiple goals. Depending on what goal the investigator has, it is perfectly legitimate for her to develop one's statistical machinery to achieve that goal. However, there is a further implication of Royall's three questions, which, Royall thinks, is that only the likelihoodist could handle the evidence question and no other school can.

6

Bayesian Accounts of Belief and Evidence Questions

In this section, I will develop two distinct Bayesian accounts to address both belief and evidence questions and argue that Royall is mistaken about his claim that Bayesians can't handle the evidence question. Although this paper disagrees with Royall, it draws its inspiration from his work regarding the nature and importance of three questions.

My two accounts to address the belief and evidence questions are an account of confirmation and an account of evidence. For Bayesians, an account of confirmation provides a confirmation relation, C(D, H,B) among data, D, hypothesis, H, and the agents' background knowledge, B. Because a confirmation relation is a belief relation, it must satisfy the probability calculus, including the rule of conditional probability together with some reasonable constraints on one's *a priori* degree of belief in an empirical proposition. Learning from experience is, of course, part and parcel of Bayesianism. The rule of conditional probability ensures this inductive basis of our learning. As a result, the agent should not have an *a priori* belief about an empirical proposition with full certainty (probability 1 or probability 0) if she would like to learn from experience, but rather something in between these two extremes.[4]

For Bayesians, then, belief is fine-grained. They allow belief to admit of any degree of strength between 0 to 1. A satisfactory Bayesian account of confirmation should be capable of capturing this notion of degree of belief. According to the confirmation condition (CC), D confirms H if and only if Prob(H|D) >0 but < 1. [5] The posterior

[4] However, I take the data in light of which the agent updates her belief function to have probability 1.

[5] One might object that this might not be a representative probabilistic measure for Bayesian confirmation. One suggestion might be to use the measure Prob(H|D) > Prob(H) instead. To respond to this objection, we would like to point out first that there are also other measures, for example, the difference between Prob(H|D and Prob(H|~D), and many more. Second, we need to distinguish between two concepts of confirmation; the *absolute* confirmation and the *incremental* confirmation (see Carnap, 1952). This notion of confirmation is intended to capture the absolute confirmation, while at the same time, making Bayesianism to be both minimally constrictive and maximally tolerant. In addition, Gill, a Bayesian sociologist, also endorses this absolute confirmation measure because, according to him, to judge the quality of a single model, there is no better measure than referring to their posterior summaries. (Gill, 2002, p. 271)

probability of H could vary between 0 and 1 exclusive. Confirmation becomes strong or weak depending on how high or low this posterior probability is.

The posterior and prior probabilities are essential parts of Bayes' theorem, a foundation of any Bayesian approach. According to the theorem, the posterior probability of a hypothesis H equals its prior probability multiplied by the likelihood of H (Prob(D|H)) then divided by the marginal probability of D (Prob(D)):

$$\text{Prob}(H \mid D) = \frac{\text{Prob}(H) \times \text{Prob}(D \mid H)}{\text{Prob}(D)} \qquad \text{(EQ 1)}$$

Prob(H|D) also is called the conditional probability of H given D. Prior probability of a hypothesis depends on the agent's degree of belief in that hypothesis before the data for the hypothesis have been gathered. The likelihood function provides an answer to the question, "How likely are the data given the hypothesis?" The marginal probability represents the probability that D would obtain, averaged over the hypotheses being true and false.

My account of evidence lays down conditions for the evidence relation to hold among D, H, and B. However, because the evidence relation is not a belief relation it need not satisfy the probability calculus. While the account of confirmation is concerned with a single hypothesis embodied in equation 1, an account of evidence must compare the merits of two competing hypotheses, H1, and H2 (or ~H1) using data D. The Bayes factor (BF) captures the bare essentials of a Bayesian account of evidence (See Raftery, 1995). For simple hypotheses, the Bayes Factor and the likelihood ratio are identical (Berger, 1985, p. 146). The BF can either be represented by a or b:

(a) D becomes E (evidence) for H1&B against H2&B if and only if

$$\left[\frac{\Pr ob(D \mid H1 \,\&\, B)}{\Pr ob(D \mid H2 \,\&\, B)}\right] > k > 1$$

(b) D provides the same evidential support (E) for both H1 & B and H2 & B if and only if

$$\left[\frac{\Pr ob(D \mid H1 \,\&\, B)}{\Pr ob(D \mid H2 \,\&\, B)}\right] = 1$$

Call these two conditions, (EQ 2a), and (EQ 2b) respectively.[6] Note that in (2a) if $1< K \leq 8$, then D is often said to provide weak evidence for H1 against H2, while when $K > 8$, D provides strong evidence (Royall, 1997). This cut-off point is determined contextually and may vary depending on the nature of the problem that confronts the investigator. The range of values for BF varies from 0 to ∞ inclusive. Note that in both (a) and (b) D represents "actual" observed data, and not "possible" unobserved data. This is the backbone of the likelihood principle. To know whether a theory is supported by data, the likelihood principle says that the only part of data that is relevant is the likelihood of the actual data given the theory. For my Bayesian account of evidence, it is the likelihood principle (LP) and not the law of likelihood that justifies the use of the Bayes Factor as a measure of evidence (Birnbaum, 1962; Berger and Wolpert, 1988, Berger, 1985, Berger and Pericchi, 1996, Good, 1983 and Rosenkrantz, 1977).[7]

The LP is derivable from two principles, the first one is called the weak sufficiency principle (WSP) and the other one is called the weak conditionality principle (WCP), which philosophers like Elliott Sober calls the principle of actualism (Sober, 1993). The

[6] There could be several objections to using the Bayes Factor as a measure of evidence. I will consider two related objections. The first objection to the Bayes Factor is that the Bayes factor could be arbitrarily large when Prob(H1) = 0. But, the objection continues, Prob(H1) = 0 means that one didn't believe that there can be any evidence for H1. In response to this objection, I could say that if Prob(H1) = 0, then the numerator in the Bayes Factor is not defined because it is of the form of the ratio 0/0, as a result, the entire Bayes factor ratio can't be evaluated. Nonetheless, if 0 < Prob(H1) < 1, and there are only two hypotheses, then BF = LR, and LR can be arbitrarily large. Large LR means strong evidence in favor of H1 over H2, but large LR does not imply large posterior probability of H1. The second objection which is a corollary to the first objection is that no account of evidence can succeed when it ignores prior probability. I have already evaluated this charge in sections 6 and 7, implying that this charge is not generally correct, because in comparing two simple hypotheses, one is not necessarily committed to invoking prior probabilities (See also Berger, 1985, p.146). However, when one is involved with a complex model, then one must appeal to the prior probability of that model. I have addressed this point in section 8. On these two points, I owe both to Robert Boik and Colin Howson for comments and suggestions. (See, Bernardo and Smith, 1994, and also Bernardo, forthcoming for their objections to the use of Bayes Factor as a measure of evidence.)

[7] Two remarks are in order. The first remark pertains to an anonymous referee's comment and its subsequent follow-up. S/he thinks that the likelihood principle does not justify the use of Bayes factor. The referee adds that there are Bayesians who criticize Bayes Factor, because they don't meet the coherence condition. What the referee says is correct. However, there are several notions of coherence in Bayesian literature. One needs to be careful about it. (For various senses of coherence, see Bandyopadhyay forthcoming). The second remark has to do with a possibility that whether one could support the LP without supporting the LL. I contend that they are different concepts, hence supporting or denying one does not necessarily lead to supporting the other and conversely. The LP only says where the information about the parameter θ is and suggests that the data summarization can be done through the likelihood function. It does not clearly indicate how to compare the evidential strength of two competing hypotheses. In contrast, the LL compares the evidential strength of two contending hypotheses. I thank both C. Andy Tsao and Jayanta Ghosh for some clarifications on the relationship between the LP and the LL.

WSP says if T is sufficient and if $T(x_1) = T(x_2)$, then both x_1 and x_2 will provide equal evidential support. A statistic is sufficient, if given it, the distribution of any other statistic does not involve the unknown parameter θ. This statistic is considered to be a desirable feature for a good estimator, and as such is embraced by both classical and Bayesian statisticians. Hence, one is ill-advised simply to reject the latter. As a result, a defense of the LP could rest on a defense of the WCP, or the principle of actualism. The WCP says that experiments not actually performed are irrelevant. According to the principle of actualism, one should make judgments about the tenability of a hypothesis based on data we actually have, rather than on unobserved, possible data. Borrowing an insight from Sober, I will justify the principle of actualism, which, indeed, will justify the LP in the end.

Consider a doctor making a diagnosis about whether a patient has diabetes (H1) or small pox (H2). There are two laboratory tests conducted, one for diabetes and the other for small pox. Their results contain both "positive" and "negative" outcomes with errors associated with these tests. Consider another test to be conducted to examine whether the patient has diabetes and this test is, in fact, infallible in the sense that it has probability one of detecting correctly if the patient suffers from diabetes. It is crucial for this defense to assume that there is no such infallible test for small pox. Suppose that the doctor runs the first two tests and not the infallible third test. Suppose each test yields a positive result with Prob(D|H1) = Prob(D|H2). Consequently, each has equal epistemological footing given actual evidence. Should we be more confident that the patient has diabetes than we should be that the patient has small pox based on our reasoning that we have an infallible test for the diabetes, but not for small pox? Our intuition about this question is "no". The principle of actualism exploits this intuition about evidence. We should base our decision relying on these *actual* results, rather than on the *possible* test results. The mere possibility of running the third test is irrelevant for the study under consideration. So far, I have defended the principle of actualism at the core of the LP, which underlies any Bayesian account of evidence.

The crucial feature for understanding the distinction between an account of evidence and that of belief confirmation is the role of the "coherence condition." Confirmation has a coherence condition (CCC) that says if H1 entails H2, then Prob(H2)

$\geq$ Prob(H1). The coherence condition is a consequence of the axioms of the probability theory. As a result, any account of confirmation must satisfy it. However, an account of evidence does not need to have a corresponding coherence condition (ECC) which says: If H1 implies H2, then whatever is evidence for H1 must also be (as good) evidence for H2.

A simple dice example will illustrate why the account of evidence does not need to satisfy this condition. Consider a pair of standard, fair dice, consisting of one red and one white die that have been thrown, but the result of the throw is unknown. The two hypotheses, H1 and H2, regarding the value of their faces would be:

(H1): the value on the faces of the two dice is equal to 3.

(H2): the value on the faces of the two dice is $\geq$ 3.

Note that H1 entails H2. Then an "evidence coherence condition" would imply that whatever is evidence for H1 must also be evidence for H2. Suppose our datum (D) is that we have learned that the red die showed 1. We have no information about the other die. Here, D favors H1 over H2, although H1 entails H2 (See Levine and Schervish, 1999 on related issues). The Bayes's factor for H1 vs. H2 is $\frac{1/2}{5/35} = \frac{1/2}{1/7} = 3.5$. So although H1 entails H2, given D, whatever is evidence for H1 need not be as strong evidence for H2, thus violating the "evidence coherence condition (ECC)." [8]

To explain the above scenario, one could say that D makes H2 makes D less likely than H1, while H1 makes D more likely. There were more ways H2 could be true than H1 before the occurrence of D. However, given H1, the occurrence of D is a bit more likely than given H2. It is important to realize that this dice example showing the violation of ECC does not rely on any prior probability or belief for either of the hypotheses.

[8] Bernardo informs me (in a private conversation) that this should be an adequate reason for rejecting the Bayes Factor as a measure of evidence, because the Bayes Factor is unable to handle evidential support involving two nested hypotheses (see also footnote 4). There is a saying among philosophers that one man's *modus ponens* is another man's *modus tollens.* As a result, contrary to Bernardo, I think that this example shows why an account of

7
Revisiting the Belief/Evidence Distinction and Four Accounts of Evidence

Continue with the PAP smear example mentioned in section I (Pagano and Gauvreau, 2000). An on-site proficiency test conducted in 1972, 1973, and 1978 evaluates the competency of technicians who examine PAP smear slides for abnormalities. I will assume from a large number of data that we are nearly certain that the propensity for having a positive PAP smear for members with cancer of the cervix is Prob(D|H) = 0.8375, and the propensity for having a positive PAP smear for population members without cervical cancer is Prob(D|~H) = 0.186. Call this background theory of the propensities for PAP smear test outcomes, “B”.

Let H represent the hypothesis that an individual is suffering from cervical cancer and ~H the hypothesis that she is not. These two hypotheses are mutually exclusive and jointly exhaustive. In addition, assume D presents a positive PAP smear test result. We would like to know Prob(H|D), i.e., the posterior probability that a person with a positive test result actually does have the disease. To apply Bayes’ theorem, we need to know the prior probability for H. Prob(H) is the probability that a woman suffers from cervical cancer when randomly selected from the population. One source reports that the rate of cases of cervical cancer among women studied in 1983-1984 was 8.3 per 100, 000 (Pagano and Gauvreau, 2000). The data yield Prob(H) = 0.000083. Then, Prob (~H) = 0.999917. By Bayes’ theorem, we get Prob(H|D) = 0.000373. Here, it tells us that for every 1,000,000 positive Pap smears, only 373 represent true cases of cancer. So, Prob(H|D) is very low; there is weak confirmation for the hypothesis. Does D provide evidence for H against ~H? The Bayes Factor-based account of evidence yields Prob(D|H)/Prob(D|~H), which equals 4.49 times. What does the value 4.49 times mean? It means that after we know that the individual has a positive result, she is five times more likely to have cervical cancer. D provides weak evidence for H against ~H. So the PAP case illustrates that the agent could have both weak belief and weak evidence for the hypothesis.

---

evidence should be different from that of confirmation, and in fact, I think, the Bayes Factor is able to capture this crucial feature of evidence.

Consider another example, which I call the Tuberculosis case (TB). The X-Ray is administered to examine whether someone is suffering from the disease. I will suppose from a large number of data that we are nearly certain that the propensity for having a positive X-ray for members with TB is 0.07333, and the propensity for a positive X-ray for population members without TB 0.0285. Call this background theory of the propensities for X-ray outcomes "B." Let H represent the hypothesis that an individual is suffering from tuberculosis and ~ H the hypothesis that she is not. These two hypotheses are mutually exclusive and jointly exhaustive. In addition, assume D represents a positive X-ray test result. We would like to find Prob(H|D), the posterior probability that an individual who tests positive for tuberculosis actually has the disease. Bayes' theorem helps to obtain that probability. However, to use the theorem, we need to know first Prob(H), Prob(~H), Prob(D|H), and Prob(D|~ H).

Prob(H) is the prior probability that an individual in the general population has tuberculosis. Because the individuals in different studies need not be chosen from the population at random, the correct frequency based prior probability of the hypothesis couldn't be obtained from them. Yet in a 1987 survey, there were 9.3 cases of tuberculosis per 100,000 population (Pagano et al, 00). Consequently, Prob(H) = 0.000093. Hence, Prob(~H) = 0.999907. Based on a large dataset kept as medical records, we are certain about these following probabilities: Prob(D|H) is the probability of a positive X-ray given that an individual has tuberculosis. Prob(D|H) = 0.7333. Prob(D|~H), the probability of a positive X-ray given that a person does not have tuberculosis, is = 1 – Prob(~ D|~ H) = 1– 0.9715. = 0.0285.

Using all this information, I compute Prob(H|D) = 0.00239. For every 100,000 positive X-rays, only 239 signal true cases of tuberculosis. The posterior probability is very low, although it is slightly higher than the prior probability. Although CC is satisfied, the hypothesis is not very well confirmed. Yet at the same time, the BF, i.e., 0.7333/0.0285 (i.e., Prob (D|H)/Prob(D|~ H)) = 25.7, is very high. Therefore, the test for tuberculosis has a great deal of evidential significance.

There is little point in denying that the meanings of "evidence" and "confirmation" (or its equivalents like "belief") often overlap in ordinary English as well as among epistemologists. There is a theorem, BF > 1 if and only if Prob(H|D) >

Prob(H), which shows this connection. However, strong belief does not imply strong evidence and the latter also does not imply the former, as illustrated by the TB example. My case for distinguishing them rests not on usage, but on the clarification in our thinking that is thus achieved and supported by inferences frequently made in diagnostic studies (See Bandyopadhyay and Brittan, forthcoming).

How good is a Bayesian account of evidence compared to Royall's likelihood-based account of evidence and Mayo's error-statistical account of evidence? Like me, Royall himself uses the likelihood ratio as a measure of evidence. Since the former and the latter is the same measure, there is no difference between these two accounts insofar as quantification of the strength of evidence of competing *simple hypotheses* is concerned. Hence, there is no difference between a Bayesian account of evidence and a likelihood account of evidence with regard to the PAP and TB cases. However, Royall and my account tend to diverge when we are confronted with a complicated problem in model selection involving a complex hypothesis (see section VIII on this point.)

Royall thinks that classical statistics or error-statistics is only able to handle the acceptance question, which is, according to him, decision theoretic. Contrary to Royall, Mayo has argued that it is a mistake to think that error-statistical account can't provide an account of evidence. I will briefly discuss Mayo's account and argue that a Bayesian account of evidence is at least as good as her error-statistical account. Mayo's error-statistical account rests heavily on the notion of severity, which she has borrowed from Karl Popper (Popper, 1959) with two major differences from the latter. First, she disagrees with Popper on the appropriate scope of falsification that underlies Popper's use of the severe test. A necessary condition for a theory to be scientific, according to Popper, is that the theory in question is in principle falsifiable. Mayo thinks that Popper is misguided in attempting to falsify a whole theory. According to her, theory testing should begin with small scale testing, such as testing a particular hypothesis with a particular outcome. Her notion of good evidence or what she calls a "severe test" is "always attached to a particular hypothesis passed or a particular inference reached" (Mayo, 1996, p. 184). The second difference between them is that while her notion of severity is probabilistic, his is deductivistic.

Consider H, and its denial, ~ H, to be mutually exclusive and jointly exhaustive of all possible hypotheses in a domain. Assume that H is a simple hypothesis, and D is the datum for H. For Mayo, H passes severely with D just in case, (i) Prob(D|H) is very high, and (ii) Prob (D| ~H) is very low. One distinct part of her account is that the notion of probability invoked for making statistical inference is frequentist/objectivist in spirit. It is qualities of experimental procedures that supply error probabilities. In her account, error-statistics rest on these error probabilities, which provide adequate information just to the extent that one should be able to make reliable statistical inferences based solely on them.

If we apply her account to the PAP and TB cases, then we find that her account considers both tests to be severe.[9] Both of her conditions for the severe test are satisfied by these two cases. In the PAP case, Prob(D|H) = 0.8375 and Prob(D|~H) =0.186, whereas in the TB case, Prob(D|H) = 0.7333, and Prob(D|~H) = 0.0285. In contrast, my Bayesian account distinguishes the strength of evidence in those two cases. In the PAP case, my account says that evidence is weak, whereas in the TB case, it yields strong evidence.

Often, the Akaikean Information Criterion (AIC) has been proposed as a measure of evidence (For Akaike's own works, see Akaike, 1973; For applications of the AIC framework in philosophy of science, see Forster and Sober 1994, Forster and Sober, 2004. See also Taper, 2004, to know more about a different slant on the possibility of arriving at better information criteria than the AIC.). The goal in using AIC is to maximize predictive accuracy and AIC provides a consistent estimate of an index of predictive accuracy. Since AIC is an estimate, the latter is computed based on both actual and possible data. Forster and Sober's recommendation is that one should choose the model with maximal AIC. The measure of evidence they recommend based on the AIC framework says that D is evidence for H1 over H2 if and only if the AIC(H1) > AIC(H2). If one uses the AIC as a measure of evidence for the PAP and TB cases, then one finds that the AIC-based measure will yield exactly the same results as given by the likelihood-based measure in those cases, because these two cases don't involve any adjustable

[9] Since Mayo is not a friend of the likelihood framework, she might object to my reducing her account to two likelihoods. There could be two responses to it. First, Mayo herself has formulated her account in terms of the likelihood (Mayo, 1996, pp.179-81). Second, if she would raise this objection to my reconstruction of her position,

parameters. Since the AIC violates the likelihood principle, like the error-statistical account it fails to be a viable account of evidence (See Boik, 2004 for more detail).

In short, my Bayesian account of evidence is at least as good as any account of evidence when the former is compared with Mayo's error-statistical account or Royall's account or for that matter with the AIC based account of evidence.

8

The Curve-Fitting problem: Accommodating Two Bayesian Accounts

So far, I have addressed only simple diagnostic examples involving a statistical hypothesis to explain the belief/evidence distinction. However, what happens when one is confronted with a complicated scenario and how do the two previous Bayesian accounts handle that case? Consider a hypothetical example to get a handle on how a complex scenario could be addressed within a Bayesian framework. Salam is a fishmonger in a small village, Shrimongal, in the northern part of Bangladesh. Two elements in Shrimongal fish-mongering are (i) torrential rains during the rainy season and (ii) lack of electrification. Salam supports his family by selling fish that he keeps under chunks of ice so that the fish will stay fresh. Now assume we have data that reflect an average monthly temperature in centigrade in Shrimongal from April to December with corresponding consumption of ice measured in cubic feet. Salam's goal is to predict how much ice he will need at different temperatures based on a given data set. To do this he needs to find a relationship, if any, between average monthly temperature and ice consumption per day. The following table **1** represents historical data for one season.

then her account would turn out be incomplete because it then fails to explain even simple diagnostic cases like the ones discussed above.

Shrimongal Data

Variable	Month								
	Apr.	May	June	Jul.	Aug.	Sept.	Oct.	Nov.	Dec.
x	Average Monthly Temperature								
	15.6	26.8	37.4	36.4	35.5	18.6	15.3	7.9	0.0
Y	Average Daily Ice Consumption in Cubic Feet								
	5.2	6.1	8.7	8.5	8.8	4.9	4.5	2.5	1.1

Table I

In the table I, the explanatory variable, x, is average monthly temperature, and the response variable, Y, is ice consumption per day. One can see that as soon as the monsoon season in June sets in, gradually temperature decreases; as a result, Salam's ice consumption goes down, hitting rock bottom in December, which is winter in Bangladesh. (This example assumes, by the way that Salam buys same amount of fish each month).

The relationship of x to Y may be represented through the following equation:

$$Y_i = \alpha_0 + \sum_{j=1}^{k} \alpha_j x_i^j + \varepsilon_i$$, for i = 1,..., n,

where n is the sample size; αj, j=0, ..., k are unknown regression coefficients, k is the order of the polynomial model, and ε_i, is random error. The error terms, ε_I, i =1,..., n are assumed to be independently distributed with mean zero and variance $\sigma 2$.

Salam wants to know what his ice consumption will be for a month this year that has x = 15.3^0 C. In October of the historical data set, 15.3 ^{0} C corresponded to 4.5 cubic feet of ice. I shall forecast from three different regression line equations how much ice Salam will require at 15.3^0 C. I consider three regression equations corresponding to three mutually exclusive hypotheses, H1, H2, or H3 in a domain. The hypotheses are,

H1: $E(Y|x) = \alpha_0 + \alpha_1 x$;
H2: $E(Y|x) = \alpha_0 + \alpha_1 x + \alpha_2 x2$; and
H3: $E(Y|x) = \alpha_0 + \alpha_1 x + \alpha_2 x2+\alpha_3 x3$.

Here, E(Y|x) is the conditional expectation of Y given x. To say that these hypotheses are mutually exclusive is to say that the coefficients of x_k under H_k are non-zero.

The least squares forecast under H1 is $\hat{Y}$ = 1.22 + 0.20 x 15.3 = 4.33 hundred cubit feet. Under H2, the prediction is $\hat{Y}$ = 1.09 +0 .22 x 15.3 -0.0005 x 15.3^2 = 4.39 hundred cubit feet per day. This is closer to the historical value of 4.5 than that based on H1. If we use H3, then we will find that the prediction, $\hat{Y}$ = 4.45 is even closer to the historical value. In general, as the order of the polynomial regression model increases from H1 to H3, the goodness of fit of the model to the observed data increases. This is measured by maximizing the likelihood function under H_k denoted by $\hat{L}_k$. That is,

$$\hat{L}_k = \max_{\sigma^2;\alpha_0,\ldots\alpha_k} Likelihood(\sigma^2;\alpha_0,\ldots,\alpha_k|H_k;Y_1,\ldots.,Y_n) \qquad \text{(EQ 3)}$$

A model having too large an order will over-fit the data. Predictions of future data from such a model will, in general, have larger errors than will predictions from a model with a smaller, but sufficient, number of parameters. So to both maximize better prediction, and minimize over-fitting error, I propose that Salam use Bayes' Theorem Criterion (BTC) to calculate how much ice he needs. The BTC implies that if one adopts certain non-informative priors on σ2 and α_j for j = 1,…,k, then the posterior probability, Prob (H_k|data), of a hypothesis is proportional to its prior probability, Prob (H_k), multiplied by the maximum likelihood function, $\hat{L}_k$ where $\hat{L}_k$ is given by (EQ 3). (For more details, see, Bandyopadhyay, Boik and Basu, 1996; and Bandyopadhyay and Boik, 1999)

I consider H1 as the simplest hypothesis because it contains the fewest number of parameters (α_0, α_1) while in light of epistemological/pragmatic considerations, it seems to be the best, and therefore I assign the highest prior probability, 1/2, to it (Our work was here influenced by Jeffreys's. However, our account of simplicity is different from his. For our work, see two immediately cited papers. For more on an account of simplicity that rests on Bayesianism, see Bandyopadhyay, 2002) Because of epistemological/pragmatic considerations, H2 gets assigned ¼ and H3 1/8. Recall that I considered only three hypotheses. The other hypotheses are lumped together. This is called the catch-all hypothesis, denoted by H_c and gets assigned 1/8. In short, BTC says, choose the one that maximizes posterior probability, that is, maximizes.

$$\text{Prob}(H_k|\text{data}) \propto \hat{L}_k \times \text{Prob}(H_k) \qquad \text{(EQ 4)}$$

The result of applying (EQ4) to the Shrimongal data yields the following table. So Salam chooses H1 as the best model for his prediction.

H1	H2	H3
-4.75	-5.30	-5.55

Table II

Applications of BTC (EQ 4) to the Shrimongal Data

Consider how this account based on the BTC is able to provide a unified account of both belief and evidence while being careful about the distinction between the two questions (Schwarz's Bayesian Information Criterion called BIC is, however, different from BTC, although we proved in the already cited papers that the former is logically equivalent to the latter with a choice of priors/substitutions. For more on the BIC see, Schwarz, 1978). Given the data about ice consumption and average monthly temperatures for a specific season, if Salam's interest can be shown to be in the evidence question, then the BF based account provides that answer. To compute the likelihood of a family under H1, and then under H2, we assign non-informative priors over regresssion co-efficients, variance and error terms. After doing this computation, what we get is that a hypothesis with higher parameters always fits the data better than a hypothesis with fewer parameters. Here, our claim just made rests on the assumption that we work within a nested model. However, if the agent opts for a hypothesis with higher dimension, then she might be involved both in an over-fitting error and the penalty due to introductions of more parameters. We need to take into account the role of simplicity in theory choice.

This leads to the curve fitting problem, which arises when one tries to optimize two conflicting desiderata, simplicity and goodness of fit. The Bayes' theorem Criterion (BTC) has been proposed to resolve the problem. An account that rests on BTC accommodates both the belief and evidence questions in a natural way and satisfies both the probability calculus and the likelihood principle that rests on using actual observed data. It could be shown that BTC satisfies the probability calculus. The former also automatically satisfies the likelihood principle which is the foundation for a Bayesian measure of evidence because the data's support for the theory stems only from the likelihood function in any Bayesian analysis (For connections between BIC and the Bayes factor see, Raftery, 1995).

9

A Decision-Theoretic Account of Acceptance

In response to Royall's acceptance question, I will provide a decision-theoretic account of acceptance. The notion of acceptance is crucial both for statistical theories as well as for theories in physical and biological sciences. Two questions arise in connection with acceptance of a theory: i) *what* is acceptance of a theory and ii) if we sometimes accept a theory, then what *justifies* this acceptance?

To answer these questions, I provide a Bayesian theory of acceptance. Building on van Fraassen's theory (van Fraassen, 1991), I defend a double aspect theory of acceptance: a) the belief aspect that states my degrees of belief in a theory and b) the pragmatic aspect that states my non-epistemic reasons for pursuing a theory, such as getting an NSF grant. In my account, I also have a justification for my double- aspect account of acceptance. A) Like Bayesians, on my view, an agent's degrees of belief must obey the probability calculus and any change in her degrees of belief must be done in accordance with the rule of conditionalization. B) As a Bayesian, I justify the agent's pragmatic reasons for pursuing a theory by invoking the principle of maximizing expected utility (hereafter, MEU).

Bayesians hold that our degrees of belief admit of a numerical representation that obeys the rules of the probability calculus. If our degrees of belief disobey the rules of probability calculus, then *some* Bayesians argue, a clever bettor can make a book against us so that we are bound to lose no matter how the world turns out to be. This argument for justifying why the agent's degree of belief should satisfy the probability calculus is known as *the Dutch-Book argument* (See, de Finetti, 1962; Skyrms, 1995. For a different perspective on the Dutch-Book argument, see Howson and Urbach, 2005). If a rational agent updates her belief in light of new data according to rules of belief change, then the rational agent's belief must satisfy the principle of conditionalization.

My Bayesian account of acceptance rests on MEU. According to MEU, in a given decision situation, the decision maker should choose the alternative with maximal expected utility. It is commonly assumed in Bayesianism that the decision maker can assign numerical values to the utilities of various outcomes in the decision situation. In Bayesianism, it is further assumed that the decision maker can assign probabilities to the states of the world. This is why Bayesian decision theory is based on the subjective probability of the event in question. For any state of the world, the decision maker can assign any probability value between 0 and 1. Although the rational agent can pick any probability value between 0 and 1, new information gathered from experience, and subject to some conditions, can change the subjective probability of the agent. Given the utility of the outcomes and the probabilities of the states, the decision maker can compute the expected utility of the various alternatives. Here, utilities are taken to be linearly related to money.

As a Bayesian I contend that in a decision situation one ought to accept the theory which has a higher expected utility than any other. Sometimes, it may happen that we accept a theory that has a lower probability than the rest of the theories in a domain, though the former has a significantly higher utility than the rest of them. As a result, based on our expected utility calculation, we end up getting a higher expected utility if we accept the theory. In contrast, we may embrace a theory that has a lower utility, but which has an appreciably higher probability than the rest of them. In the end, we embrace the former, because we obtain a higher expected utility if we embrace it.

Summing up: Dreams of a Final Theory

The purpose of this paper was to bring Royall's three questions to a wider audience and assess them from a Bayesian perspective. In addition, the paper also provided analyses for other topics. I discussed both Bayesian and non-Bayesian considerations for imposing *a priori* desiderata on inductive/statistical inference. I diagnosed that the choice of desiderata depends on one's goals, which could be multiple. I discussed why Royall's emphasis of multiple goals in terms of three questions is rewarding, although I pointed out a defect in the law of likelihood on which his likelihood framework rests. I also defended the likelihood principle. I developed two distinct accounts within Bayesianism to respond to the belief and evidence questions and discussed how my Bayesian account could tackle complicated problems that often arise in model selection. Based on the idea of maximizing expected utility, I have provided a response to Royall's acceptance question. My Bayesian answer to the acceptance question is that we should choose the theory that has a higher expected utility than its rival.

In the literature on scientific explanation, an explanation is considered to be a good scientific explanation if it is able to unify diverse phenomena. A theory that has the ability to provide a unifying explanation for several apparently unrelated phenomena is hailed as a good theory. For example, we prefer Einstein's theory of relativity to Newton's theory because of this reason. Using Einstein's theory, we could explain (i) the occurrence of a red-shift, (ii) the bending of light in front of massive nearby objects and finally (iii) the precession rate of Mercury's perihelion with sufficiently precise details. Although the ability to unify diverse items under one banner is canvassed as a plus point for scientific explanation, I think that this ability should also be counted as an added advantage for a statistical account that provides a unifying view of several apparently disjoint questions. I argued that Bayesianism and Bayesianism alone is able to provide an unified account to all three questions thus providing a primer for philosophers of science interested in a unified approach to issues like belief, evidence, statistical inference and decisions about which theories they should accept.

Select Bibliography

1. Adams, E. 1998. *A Primer of Probability Logic*. Stanford, CA:CSLI
2. Akaike, H. 1973. Information Theory as an Extension of the Maximum Likelihood principle. In Petrov, B.N., and Csaki, F. ed., *Second International Symposium on Information Theory*. Budapest: Akademia Kiado.
3. Bandyopadhyay, P.S. Types of Coherence and Coherence among Types. In Pereira, L. M., and Wheeler, G. (eds.). In *Computational Models of Scientific Reasoning and Applications*: CMSRA-IV, Lisbon, 2005: 129-143.
4. Bandyopadhyay, P.S., and Brittan, G. Jr. Acceptance, Severity and Evidence. *Synthese* (forthcoming): 148 (2). 259-293.
5. Bandyopadhyay, P.S., and Bennett, J. G. 2004. Commentary on Mauer. In Taper, M.L., and Lele, S.R. eds., *The Nature of Scientific Evidence*. Chicago: University of Chicago Press, 32-39.
6. Bandyopadhyay, P.S. 2002. Simplicity: Our View, Their View. Presented at the *American Philosophical Association*. Chicago: Central Divisional meetings.
7. Bandyopadhyay, P.S., and Boik, R.J. 1999. The Curve Fitting Problem: A Bayesian Rejoinder. *Philosophy of Science*, 66 (supplement): 391-402.
8. Bandyopadhyay, P.S., and Boik, R. J. and Basu, P. 1996. The Curve Fitting Problem: A Bayesian Approach. *Philosophy of Science*, 63 (supplement): 264-272.
9. Berger, J. O. 2000. Bayesian Analysis: A Look at Today and Thoughts of Tomorrow. *Journal of the American Statistical Association*, 95: 1269-1276.
10. Berger, J. O. 1985. *Statistical Decision Theory and Bayesian Analysis*. Second edition, New York: Springer.
11. Berger, J.O., and Wolpert, R. L. 1988. *The Likelihood Principle*. Hayward, CA: Institute of Mathematical Statistics.
12. Berger, J.O., and Pericchi, L. 1996. The Intrinsic Bayes Factor for Model Selection and Prediction. *Journal of the American Statistical Association*, 91:pp. 109-122.
13. Bernardo, J.M., and Smith, A. F.M. 2000. *Bayesian Theory*. Weinheim: John Wiley.
14. Bernardo, J.M. (forthcoming). Reference Analysis, p.1-86. In *Handbook of Bayesian Statistics*. Dey, D. and Rao, C.R. (eds.) Elsevier.
15. Birnbaum, A. 1962. On the Foundations of Statistical Inference (with discussion.) *Journal of the American Statistical Association* 57: 269-326.
16. Boik, R. J. 2004. Commentary on Forster and Sober. In Taper, M.L., and Lele, S.R. eds., *The Nature of Scientific Evidence*. Chicago: University of Chicago Press.
17. Burnham, K. P., and Anderson, D.R. 1998. *Model Selection and Inference* New York: Springer.
18. Carnap, R.1952. *The Continuum of Inductive Methods*. Macmillan, New York.
19. Cassella, G., and Berger, R.L. 1990. *Statistical Inference*. Belmont, Ca: Duxbury.
20. Christensen, D. 1999. Measuring Confirmation. In *Journal of Philosophy*. 99, No 9, Sept, 437-461
21. Dawid, A. P., and Morera, J. 1996. Coherent Analysis of Forensic Identification of Evidence. In *Journal of the Royal Statistical Society.* Series B (Methodological), Vol, 58, No 2, 425-443.
22. de Finetti, B.1937. La prevision: ses lois logiques, ses sources subjectives. English trans., "Foresight: Its Logical Laws, Its Subjective Sources." In Kyburg, H.E. Jr., and Smokler, H.E. eds., *Studies in Subjective Probability*. Huntington, NY: Kreiger, 1962.
23. Descartes, R. 1911. *The Philosophical Works of Descartes*, eds. Haldane, E., and Ross, G.R.T. Cambridge: Cambridge University Press.

24. Fitelson, B. 1999. The Plurality of Bayesian Measures of Confirmation and the Problem of Measure Sensitivity. *Philosophy of Science*, 66 (supplement): 362-378.
25. Forster, M., and Sober, E. 2004. Why Likelihood? Chapter 6. In Taper, M.L., and. Lele, S.R eds., *The Nature of Scientific Evidence*. Chicago: University of Chicago Press.
26. Forster, M., and Sober, E. 1994. How to Tell When Simpler, More Unified, or Less *Ad Hoc* Theories Will Provide More Accurate Predictions. *British Journal for the Philosophy of Science*. 45:1-35.
27. Ghosh, J.K., ed. 1988. *Statistical Information and Likelihood*. New York: Springer.
28. Gill, J. 2002. *Bayesian Methods*. London. Chapman & Hall
29. Good, I. J. 1983. *Good Thinking*. Minneapolis: University of Minnesota Press.
30. Howson, C., and Urbach, P. 2005. *Scientific Reasoning*: *The Bayesian Approach*. Third edition, La Salle, IL: Open Court.
31. Jaynes, E. 2003. *Probability Theory: The Logic of Science*. Cambridge: Cambridge University Press.
32. Jeffreys, H. 1957. *Scientific Inference*. Second edition. Cambridge: Cambridge University Press.
33. Jeffreys, H. 1961. *Theory of Probability*. Third edition. Oxford: Clarendon Press.
34. Leibniz, G. *Philosophical Schriften von G.W. Leibniz,* ed. Gerhardt, C. I. Berlin, 1875-90, vol. IV.
35. Lele, S. R. 2004. Evidence Functions and the Optimality of the Law of Likelihood. In Taper, M.L., and Lele, S.R. eds., *The Nature of Scientific Evidence*. Chicago: University of Chicago Press.
36. Levine, M. and Schervish, M. J. 1999. Bayes Factor: What They Are and What They Are not. *The American Statistician*, Vol. 53, No 2 (May), 119-122.
37. Mayo, D. 2004. An Error-Statistical Philosophy of Evidence. In Taper, M.L., and Lele, S.R. eds. *The Nature of Scientific Evidence*. Chicago: University of Chicago Press.
38. Mayo, D. 1996. *Error and the Growth of Experimental Knowledge*, University of Chicago Press: Chicago.
39. Neyman, J. 1967. *A Selection of Early Statistical Papers of J. Neyman*. Berkeley: University of California Press.
40. Pagano, M. and Gauvreau, K. 2000. *Principles of BioStatistics*. Duxbury, Australia.
41. Popper, K. R. 1959.*The Logic of Scientific Discovery*. New York: Basic Books.
42. Raftery, A. 1995. Bayesian Model Selection in Social Research (with discussion). In *Sociological Methodology 1995*, ed. Marsden, P. V.
43. Rosenkrantz, R. D. 1977. *Inference, Method, and Decision*. Dordrecht: Reidel.
44. Royall, R. 1997. *Statistical Evidence: A Likelihood Paradigm*. New York, Chapman Hall.
45. Schwarz, G. 1978. Estimating the Dimension of a Model. *Annals of Statistics*. 6: pp.461-464.
46. Sober, E. 1993. Epistemology for Empiricists. In *Midwest Studies in Philosophy*, XVIII (eds.) French, P., Uehling, T. and Weinstein, H.
47. Sober, E. 2005. Is Drift a Serious Alternative to Natural Selection as Explanation of Complex Adaptive Traits. In O'Hear, A. (ed.): *Philosophy, Biology and Life*. Cambridge: Cambridge University Press.
48. Skyrms, B. 1995. Strict Coherence, Sigma Coherence and the Metaphysics of Quantity. *Philosophical Studies*. January issue.
49. Taper, M. L. 2004. Model Identification from Many Candidates. In Taper, M.L., and Lele, S.R. eds., *The Nature of Scientific Evidence*. Chicago: University of Chicago Press.

50. Taper, M.L and Lele, S.R. 2004. (eds.), *The Nature of Scientific Evidence*. Chicago: University of Chicago Press.
51. Van Fraassen, B. C. 1991. *Quantum Mechanics: An Empiricist View*. Oxford: Clarendon Press.
52. Weinberg, S. 1992. *Dreams of a Final Theory*. New York: Pantheon Books.
53. Zellner, A. 1997. *Bayesian Analysis in Econometrics and Statistics*. Cheltenham, UK: Edward Elgar Publishing Limited.
54. Zellner, A. 1996. *An Introduction to Bayesian Inference in Econometrics*. New York: John Wiley.
55. Zellner, A. 1984. *Basic Issues in Econometrics*. Chicago: University of Chicago Press.
56. Zellner, A.2001. Keep it Sophistically Simple. In A. Zellner., Keuzenkamp, H. and McAleer, M. eds., *Simplicity, Inference and Modelling*. Cambridge: Cambridge University Press.