The Scope of Bayesian Reasoning[1]

Henry Kyburg, Jr.

University of Rochester

1. One View of Bayes' Theorem

There is one sense in which Bayes' theorem, and its use in statistics and in scientific inference, is clearly uncontroversial. It is an authentic, certified, theorem of the probability calculus, and even the founders of classical statistical inference, Fisher, Neyman and Pearson, were explicit about seeing no difficulty in the use of Bayes' theorem when the conditions for its application were satisfied. For example, Fisher writes, "When there really is exact knowledge *a priori* Bayes' method is available" (1971, p. 194).

What are these conditions? Why, simply that a joint distribution be known that supports the inference from a sample distribution to a posterior distribution for the hypotheses in question. Let me give a very brief example of a context in which everyone would seem to be in happy agreement, though their descriptions would vary as a function of their views of probability.

We have an experiment in which we choose one of two urns, urn-1 and urn-2, with each having equal chance of being chosen, and then choose a ball from the urn, each ball having the same chance of being chosen. Urn-1 contains two balls, one white and one black; urn-2 contains three balls, one white and two black. Clearly the chosen ball gives us some knowledge about which urn we have chosen, if we don't already know.

The joint distribution can be computed easily enough: $p(1\&B)=p(1\&W) = 1/4$; $p(2\&B) = 2/6$; $p(2\&W) = 1/6$. The prior probability associated with each urn is a half. When we draw a black ball, the conditional probabilities become $p(1|B) = p(1\&B)/(p(1\&B) + p(2\&B)) = (1/4)/(1/4 + 2/6) = 3/7$, and $p(2|B) = 4/7$. These probabilities represent the posterior distribution.

All this may hold however you construe probability. A frequentist will say that what we are describing are the long run properties of a repeatable experiment: 3/7ths of the time, when you do an experiment of this sort and get a black ball, you will have chosen urn-1. A logical theorist will say that in the language in which the experiment

---

has been described, the appropriate measures on the sentences are such that the conditional logical probability of urn-1, given a black ball, is 3/7ths. A subjectivist will say that my opinions made coherent yield this measure.

It is important to see that in the example I have just described, so far as I know, everyone will agree that the prior probabilities exist, that the posterior probabilities have the values I attribute to them, and that the mechanism for getting to the posterior probabilities is Bayes' theorem.

What is controversial about this example is whether the probability is to be attributed only to the class of trials (actual or hypothetical) of this experiment, or whether it makes sense to attribute the probability to having chosen urn-1 on a particular occasion—say the trial of this experiment occurring at 11:00 AM on Friday, October 1, 1992. The serious frequentist, as I interpret him, will deny the latter possibility: probability makes sense only when attributed to general classes or properties. This is a view that, in common with Colin Howson and Peter Urbach (1989), I think mistaken; it leads to a variety of difficulties that have been noted repeatedly in the literature on the foundations of statistics, particularly by writers of Bayesian persuasion.

It is worth noting, however, that even from this (mistaken) point of view, the application of Bayes' theorem can be generalized to some degree. Let me begin by stating the classical form a bit more generally: We have a probability distribution over a space consisting of a number of hypotheses (the two urns in our first example) and outcomes of experiments (drawing a ball and noting the color in that example). Given the outcome O_j of an experiment, we compute the probability of one of these hypotheses as follows:

$$P(H_i|O_j) \ = \ P(H_i)P(O_j|H_i) \, /P(O_j),$$

where $P(O_j)$ can be expanded as

$$P(O_j) \ = \ \Sigma \ P(H_k)P(O_j|H_k),$$

the summation extending over all the hypotheses.

In many cases we may not know the prior distribution over the hypotheses exactly but nevertheless be willing to put constraints on that distribution. In our example, we may not be willing to say that the chance of picking each urn is exactly a half, but only that (say) it is at least 0.2 for each urn. Initially, that is to say, we do not endorse a single point valued distribution, but a set of them, **P**, which represents what we take ourselves to know about the experiment. Note that this is not clasically "Bayesian" since we are employing a set of distributions rather than a single distribution. We can still use Bayes' theorem however.

As the result of conditionalization, we will not get a single distribution, but a new *set of* distributions. If the original joint distribution is the family **P**(H,O), then the prior distribution for the hypotheses H is the marginalization,

$$\mathbf{P}(H) \ = \ \{P(H): P(H) = \Sigma \, Q(H\&O) \ \& \ Q \in \ \mathbf{P}\},$$

where the summation extends over all O consistent with H, and the new family indicated by the evidence is

$$\mathbf{P}(H|O) \ = \ \{P(H|O): P \in \ \mathbf{P} \ \& \ P(O) \neq 0 \}.$$

It is worth mentioning this natural and simple extension of the relatively uncontroversial form of Bayes theorem for two reasons: First, some classical statisticians, for example Fisher, do not regard it as legitimate. Fisher claims that Bayes theorem can be applied only when you have an *exact* prior distribution. Second, and more important, is the fact that so-called *robust* inference of this form makes it easier to believe that Bayesian inference can be extended more widely in scientific reasoning than some conservatives might think.

I leave aside the question here of the structure of the set **P**. Some writers, for example Levi (1974), suggest that the set should be convex. It has been argued (Kyburg and Pittarelli 1992) that this constraint leads to difficulties, in view of the fact that the convex combination of distributions embodying independence need no longer exhibit independence.

2. Prior Probabilities

We noted above that in a certain sense the move to sets of distributions as the input for Bayes' theorem is not really *Bayesian* in the programmatic sense of the word. In the view of most logical or subjectivistic Bayesians, what the individual should start with is a single coherent probability distribution, though logical shortcomings may make this difficult. As Howson points out, it is exactly this that ensures *consistency* in the sense that the set of fair odds representing the individuals beliefs is really *fair*. I propose now to examine the feasibility and plausibility of an assignment of probability to sentences or propositions in this classical sense: that an individual has exactly one such probability distribution. But let us keep in mind the application of Bayes' theorem to sets of probability distributions for two reasons: First, it is often desirable to represent the opinions of groups of individuals; and second, it may be an option that alleviates the difficulty of pinpointing degrees of belief for an individual.

There are a number of ways of thinking of the assignment of a priori probabilities. They may be construed as subjective ; they may be construed as logical measures on the sentences of a formal language; they may be construed as logical measures on the sets of worlds corresponding to propositions; they may be construed as relative to a set of answers to a question or problem, as I take it the maximum entropy approach proposes. And of course, as I noted earlier, they may be taken to be solidly based on our knowledge of frequencies or chances in the actual world.

I will assume that the objects to which we assign probabilities are sentences of a formal first order language. This language may mirror a fragment of ordinary English, so that you can think of probabilities as being assigned to sentences in English, if you prefer.

The first problem we face is that if the language purports to be at all global—or to be the factual fragment of English—there are a great many sentences—surely a denumerable number. While a formal language may be restricted to embodying a finite number of logically distinct sentences, such a language can hardly interest us in the general context of scientific reasoning. To avoid focussing on the peculiarities of a particular language, let us focus on the models of that language. There are, then, a denumerable number of distinct models in the intended interpretation of the language with which we are concerned.

Our first problem is that there seems to be no feasible way in which to assign probabilities to those models. Of course we can assign probabilities to certain *sets* of those models. I assign the probability 1/2 to the set of models of L in which the sen-

tence "the next toss of a coin I perform will come up heads" is true. But the general view requires us to be able to assign measures to any sentence at all, and this clearly requires that we assign measures to the individual models of our language.

Now it would not be reasonable to demand of someone that he or she make a denumerable number of specifications all at once. That would be hard work, even for the physically fit. But one should be able to approach this. But as Gilbert Harman (1986) argues, it is hard to do this even in very simple and artificial cases. "If one is to be prepared for various possible conditionalizations, then for every proposition P one wants to update, one must already have assigned probabilities to various conjunctions of P together with one or more of the possible evidence propositions and/or their denials. Unhappily this leads to a combinatorial explosion, since the number of such conjunctions is an exponential function of the number of possibly relevant evidence propositions. ... For thirty evidence propositions, a billion probabilities are needed, and so on" (p. 26). Even in a limited way, the direct approach seems not feasible, even leaving to one side the difficulty of ensuring that the assignments are consistent.

It is clear, then, why many writers have opted for systematic assignments of probability to the models of a language (or to sets of models). The classical views of Carnap (1950), Hintikka (1966), and others provide for the assignment of probabilities to the sentences of a language based on a canonical procedure. As Howson points out, such procedures are not without arbitrariness. In particular, the richer such a language is taken to be, the more parameters are involved in characterizing the "logical measure function," and the more apparent it is that some kind of personal judgment is playing a role.

It is playing a role in two distinct ways. One is in the selection of the values of the parameters that will go to generate the measure function. The other is in the selection of the language itself. This is a feature of any theory according to which the sentences of a language can bear probabilities. It is obscured by simply writing in one's native tongue as though that were not a language, but reflection reveals that it is, after all, the sentences of that language whose probabilities one is discussing.

Another approach is to look at matters more locally. Harman's argument suggests that it is implausible, even in a very limited local context, to assign probabilities purely arbitrarily, but there are suggestions according to which we can assign probabilities systematically in limited contexts. One such is the suggestion of E. T. Jaynes (1958) that we should assign prior probabilities in such a way as to minimize information (or to maximize entropy). Again, as Howson points out, this *is* an assignment of probability, and one which is arbitrary in the sense that another might have been made. It is not forced on us.

There is another consideration. Many subjectivists find the principle of countable additivity—the principle that the probability to be assigned to a countable union of exclusive propositions should be the countable sum of the probabilities assigned to the individual propositions—unacceptable. Given a countable number of exclusive alternatives, then, they will insist that no more than a finite number can receive positive (bounded by δ) probability. Applied to the models of the language, that means that no more than a finite number of models may carry bounded probability. This does not answer Harman, since the finite number can get very large very fast, but it does at least provide an "in principle" argument: in principle a finite number of assignments will suffice.

Unfortunately, this solution clashes with another subjectivistic principle The subjectivist (Colin Howson, to pick a non-random example) argues against accep-

tance—against assigning full belief, probability 1, to any non-datum sentence. (We'll worry about data later.) In particular, he argues that it is absurd to suppose that we "accept" the result of a statistical test because that means we would be assigning a probability of one to it, and *surely* we must allow for the possibility of being wrong. In general, the argument against an inductive logic that leads to the acceptance of hypotheses, as distinct from one which assigns probabilities to hypotheses, is exactly that we should never assign a probability of one to a hypothesis that *might* be wrong.

If we assign positive probability to only finitely many models, we must assign 0 probability to each of the denumerable remainder, and thus to every proposition that may be identified with a set of these models. But to assign 0 to a proposition is to assign 1 to its denial. This clearly conflicts with the injunction to eschew "acceptance" or the assignment of probability 1 to contingent statements.

3. Direct Inference

Direct inference is the principle that allows you to pass from knowledge of a chance or frequency of a property (half the tosses land heads; the chance of a head is a half) to the probability that a specific instance (the next toss, the last toss) will have that property. Obviously this principle must be hedged around with conditions in order to be applied with consistent results. For example, suppose that Tom is a miner and a Baptist; we know that the chance that a miner survives for a year is .917; we know that the frequency with which Baptist miners survive for a year is .950. We cannot have the probability that Tom survives for a year be both .917 and .950, though Tom is an instance of each of the reference classes mentioned, or alternatively is subject to both chances. We must adopt some conditions that will allow us to use our knowledge of chances and frequencies consistently.

In the classical tradition of the early twentieth century, direct inference was the inference to the probability distribution of characteristics of a sample, from the statistical premise that gave the distribution in the population. For example, from the premise that the characteristic function of *heads* is binomially distributed in the set of coin-tosses, we may infer that X_n, the number of heads on n tosses, is approximately normally distributed with a mean of a half.

Direct inference was contrasted with "inverse inference,"[2] which was regarded as suspect, and involved the inductive inference from the characteristics of a sample to the characteristics of the population from which the sample was drawn. For example, to examine an initial segment of a sequence of coin tosses, and infer something about the distribution of *heads* in the whole sequence. Bayes' theorem would allow us to do this if we had a prior distribution over the distributions that *heads* might have. But where could this come from? Or how can we apply Bayes' theorem without it? In the early part of this century, statisticians wrestled with this problem of "inverse inference"—a combat from which R. A. Fisher (1924, 1930) and then Neyman and Pearson (1928) rescued them by arguing that inverse inference was unnecessary. Recently we are being told that inverse inference is the right way to go after all, for that is just the Bayesian Doctrine.

Direct inference has seemed relatively uncontroversial until recently. Since 1959 I have argued that direct inference, though more complicated than people have thought, is all we need. Carnap (1971), more recently, has taken it to represent an important principle. David Miller discussed the principle in 1967, and argued that in a Carnapian framework it leads to inconsistency. David Lewis (1980)has baptised it the "Principal Principle" and argued that it is the glue that ties objective probability and subjective probability together. Howson claims that this principle is central to Scientific

Reasoning, and that it would be a "disaster" if it were, as Miller claims, inconsistent.[3]

Lewis's formulation, like that of David Miller, can be put this way:

P(Fa|Ga&Chance(F,G) = r) = r.

Let us call this the *stark* version of the principle. Stated thus the principle is essentially vacuous. It is quite true that if "all I know" is that a is G, and that the chance (alternatively, frequency) of a G being an F is r, then the probability for me that a is F should be r. But of course that is not "all I know" and can't be "all I know."

While it may be logically possible that my corpus of knowledge contains exactly "Ga&Chance(F,G) = r," it is surely not epistemically possible. Even if it were to be epistemically possible, it would not apply to us. We know, always, a lot more than that. In order for the principle to serve its purpose, it must be expressed thus:

P(Fa|Ga&Chance(F,G) = r&K) = r,

where K represents the other stuff that we know. Stated thus, it becomes clear that we need a proviso: that K not contain anything relevant to "Fa," other than "Ga&Chance(F,G) = r." To spell out what this means is exactly to spell out criteria for the choice of a reference class or the choice of a chance set-up, or general epistemic criteria of relevance, or something analogous.

To see this, we need merely note that the constant "a" in the principle is generally instantiated by a definite description (the next toss of the coin, the next sample of n to be chosen, the result of the coin toss performed at (time,place), ... A proper name gives us no handle, unless we have a definite description to single out its referent. But as soon as we have a definite description, we have a lot of information that must be taken account of.

To spell out conditions of relevance is, as those of us who have been working on the problem know only too well, very difficult. A complete discussion of direct inference would not be appropriate here. But it will be illustrative to exhibit several constraints on direct inference that will show how non-trivial these constraints are.

(1) Suppose that we know that a belongs to B, and to $B \cap C$, and that the proportion of B's that are T is .3, and the proportion of $B \cap C$ that are T is .6. Clearly the appropriate probability, other things being equal, is 0.6. This is entailed by Hans Reichenbach's principle: always select the narrowest reference class about which you have statistics.

(2) Suppose that we know that a is selected from B, which in turn is selected from **B** and that for every B_i in **B**, the proportion of T's is p_i, and that there are n members of **B**. We may also know that the frequency of T's among the whole union of B_i's is q. It is clear that $1/n$ times the sum of the p_i is to be preferred to q.

(3) Suppose the proportion of black balls in an urn is known to be p, but that we have selected a large number of balls from the urn, and have good reason to believe that the long run frequency of black balls among balls *selected* is q, rather than p. Clearly q is to be preferred.

(4) Consider the hypothesis H that 20% of the draws of balls from an urn yield a black ball. We take a sample of draws, and 22% are black. Relative to this information, the probability of H may be quite high. Now we continue our sampling.

Of the total sample, we find that 30% are black. Relative to this information the probability of H may be quite low. Clearly the second probability is the one to be preferred, even though our original evidence is still part of our body of knowledge.

These are the sorts of problems that make the formulation of a consistent principle of direct inference difficult. They are avoided by stating the principle in relation to a body of evidence that contains *only* one statistical or chance statement, and a statement to the effect that a given individual belongs to the reference class the chance statement concerns. What has been common in the literature is to pass from the plausible defense of the stark principle of direct inference to the mushy "if there is nothing else in the body of knowledge that bears on the result..."

But this transition is exactly what makes the selection of a reference class difficult. It is exactly what calls for careful and thoughtful analysis.

4. Subjectivity: Convergence

The most common complaint about Subjective Bayesianism is that it is subjective. There are three general responses to this charge, which we shall consider in turn. The first is that the subjectivity involved becomes diminished as evidence accumulates; the second is that subjectivity infects everything anyway; and the third is that although the input to Bayesian inference is subjective, the process of inference itself is perfectly objective.

It was de Finetti (1937) who first made subjective Bayesianism statistically respectable by showing that opinions converge as evidence mounts. What was shown originally was that if you have an exchangeable sequence of events, each of which has or lacks a property P, two people with differing non-extreme opinions about the probability that the next event will have P, will differ less and less as they condition their beliefs on a longer and longer initial segment of the sequence. "Non-extreme" opinions are those which (i) assign a probability other than 0 or 1 to P, and (ii) do *not* assume that the events are independent (else conditionalization would not change the original belief state).

In addition it is required that the sequence be *exchangeable* with respect to P, according to both parties. For the sequence to be exchangeable according to an opinion, is for the probability of any sequence of n occurrences of P and not-P to depend only on the number of P's and the number of not-P's. Example: if Heads is exchangeable in a sequence of coin tosses, *HHHHTTTT* will have the same probability as *HHTHTTHT*.

More generally, if a sequence is *exchangeable* with respect to a random quantity Q—a function that takes on a numerical value for each member of the sequence—according to two non-extreme subjective opinions, then as these opinions are conditioned on a longer and longer initial segment of the the sequence, they will come to be closer and closer together. This result can be generalized in yet further ways, so that the sequence need not be fully exchangeable, but only partially exchangeable.

This result is used to argue that subjectivity in initial opinions is unimportant because differences of opinion will be wiped out by increasing evidence. There are a number of gaps between the premise and the conclusion.

The theoretical results concern sequences of a special sort: not opinions in general. To support the argument that differences of opinion are unimportant, we would need to be convinced that all differences of opinion, and not just those concerning ex-

changeable sequences, will tend to be reduced by the accumulation of evidence. It is not at all clear that this is the case.

The technical results require that the two opinions whose convergence we are concerned about both agree that the sequence in question is exchangeable: fifty heads followed by fifty tails must be exactly as probable as any other order of fifty heads and fifty tails.

While it is not hard to agree that stubborn opinions that assign 0 or 1 to a proposition are going to be hard to alter (but recall section 3 in which we showed that *most* opinions must be 0 or 1), it is not so clear that judgments of independence are to be eschewed: if we are to learn from experience by conditionalization, we are prohibited from supposing that the outcomes of two coin tosses are independent. But many of us would be surer about this than about a lot of other things.

Let us assume that it is true, though it is difficult to see how it could be shown, that opinions in general converge with increasing evidence. More precisely, let us suppose that it is true that for any δ and any proposition S, if neither P_1 (opinion 1) nor P_2 (opinion 2) assigns 0 or 1 to the probability of S, then there is some body of evidence E, neither entailing S nor entailing $-S$, such that $|P_1(S|E) - P_2(S|E)| < \delta$. Does this remove the sting of subjectivity?

No, because I must agree with my friend on a course of action right now. We cannot wait to acquire enough evidence that our conditional probabilities are in agreement to the extent required to yield the same decision. Convergence in the indefinite future does not assuage the difficulties of our subjective differences now. In the long run, as Keynes said... .

Furthermore, it is easy to change the order of quantification, and say, with equal conviction, that for any Δ and any proposition S, and any body of evidence E, there exist prior opinions P_1 and P_2 such that neither is extreme, and yet such that $|P_1(S|E) - P_2(S|E)| > \Delta$. To see this, simply note that

$$P_1(S|E) = P_1(S)/(P_1(S) + k_1(1-P_1(S))) \text{ and}$$

$$P_2(S|E) = P_2(S)/(P_2(S) + k_2(1-P_2(S))).$$

The constants k_1 and k_2 represent likelihood ratios: $P_i(E|S)/P_i(E|\sim S)$. These ratios can have any value between 0 and infinity, and thus the difference in the conditional probabilities can be made larger than Δ. The convergence arguments do not seem to carry as much weight as some people suppose.

5. Subjectivity: Pervasiveness

The second sort of argument in defense of the subjectivism of subjective Bayesianism claims that subjectivism infects any other approach as well. For example, in the theory of testing statistical hypotheses, the choice of a particular test, or of a particular test level, may be seen as arbitrary. It has been argued (for example by Howson, p. 196) that no objective defense of a shortest confidence interval is possible, since what is shortest for t need not be shortest for $f(t)$. In short, it is claimed that subjectivism is inevitable. (p.289) "No prior distribution reflects only factual data unmixed with anybody's opinions." To the suggestion that some assumptions are gratuitous (e.g., that the laws of nature are wildly different in remote parts of the universe), while others are not, Howson replies (p. 289) "... *any* assumption imports knowledge."

It is difficult to approach the question of whether or not subjectivism is inevitable in a cool and calm manner, since Subjective Relativity in All Things seems to be the current politically correct watchword. But here we are concerned only about science, and there is certainly a common feeling that science, at least, is objective: science should follow where the evidence points, and be independent of political, moral, and subjective constraints.

Does scientific inference contain an irreducible and inevitable subjectivistic element? This is not an easy question to answer. Consider statistics. Savage showed (1962) that in deciding between two simple hypotheses, the choice of power and size of a statistical test corresponded exactly to choosing a prior probability. As Howson convincingly argues (pp 189ff) much of classical statistical inference is subject to many of the same complaints that Bayesian statistics is subject to. But the state of current statistical theory does not present a picture of clarity and agreement. There are many controversies in the foundations of statistics that are unresolved right now.

Here is just one example. Suppose you know that the quantity Q is distributed normally, with an unknown mean μ, and a known variance $\sigma^2 = 1.0$. Draw a sample of one, and observe the value x of Q. Since $\mu - x$ has a known distribution—it is Normal(0,1)—we can simply look up in a table the probability that $|\mu - x|$ exceeds any given amount. Thus if we observe that $x = 10$, we can (by careful direct inference!) conclude that the probability that $9 \leq \mu \leq 11$ is 0.68. We have used statistical knowledge, of course—the knowledge concerning the distribution of Q—but it is not at all clear that we have used any knowledge concerning the prior distribution of μ.

This is, in fact, an illustration of what R. A. Fisher called 'fiducial inference.' It has been discussed by Bayesian statisticians, who claim that there *is* a prior distribution of μ taken for granted, namely, a uniform distribution. The reason that the use of this prior distribution has escaped the attention of some of us is that it is the *improper uniform distribution* that takes every interval of equal size of possible values for μ to be equally probable a priori. And sure enough, if you take that as the prior probability distribution for μ, and perform a Bayesian analysis, you get the same results.

But does this really show that a prior distribution is taken for granted in this piece of statistical inference? Not to my way of thinking, though of course it opens up that possibility. There is no reason, in this example, to introduce a prior distribution at all. The inference can perfectly well be constructed as a simple case of direct inference, in which case it is not clear where the "subjective" element enters in. It is clearly not in the assumption of the normality of the distribution of Q, or its variance, since we assumed those to be objective facts. These assumptions could be wrong, of course., but they purport to be objective. That is another question, and does not undermine their objectivity of these alleged facts.

Note, in fact, that this is almost a touchstone of objectivity: the possibility of error. There is no way I can be in error in my prior distribution for μ—unless I make a logical error—whether I take it to be the improper uniform prior or any other coherent prior. It is that very fact that makes this prior distribution perniciously subjective. It represents an assumption that has consequences, but cannot be corrected by criticism or further evidence.

6. Subjectivy: Inference

The third defense against charges that Bayesianism embodies too much subjectivity is (so far as I know) unique to Colin Howson. It is that there is *no* subjectivity in

the Bayesian approach. (This does seem to constitute a pragmatic contradiction of the defense, also offered by Howson, that everyone else is subjective, too.) Howson speaks of the "constraints imposed on [probabilities] by the condition of consistency," and says "... there is nothing subjective in the Bayesian theory *as a theory of inference*: its canons of inductive reasoning are quite impartial and objective." (p. 296)

As a theory of inference, the calculus of probability—which is what embodies the "canons of inductive reasoning" on the Bayesian view—is, like any other piece of mathematics, a purely deductive system. It thus is surely objective, and embodies nothing controversial. It is the role it is to play in scientific reasoning that matters.

Howson takes classical statistics to task for leaping to conclusions, in, for example, rejecting a null hypothesis on the basis given evidence, on the argument that such rejections will rarely be mistaken. "... we regard such inductions as unwarranted, and the supporting argument as fallacious." (p. 190) That is because no conclusion inferred from a statistical test, no confidence interval, is immune from further testing, or from retraction in the face of new evidence.

Of course this is just to say that conclusions about matters of fact are corrigible. It is not to say that there may not be computational and epistemic advantages to *accepting* such conclusions. This is eminently clear in the writings of Fisher, who regarded hypothesis testing as a preliminary stage of scientific investigation. First we reject the null hypothesis that the treatment has no effect; and then we buckle down to work in our laboratory (or our fields) to discover what the effect is and how it is produced. We do so, however, fully recognizing that our initial rejection may have been wrong. This hardly conforms to the caricature of classical statistics according to which the rejection of a hypothesis is (or ought to be) eternal.

It is perfectly consistent to take "Inductive logic [to be] ... the theory of inference from some exogenous given data and prior distribution of belief to a posterior distribution." (p. 290) It does leave us with the puzzling Bayesian treatment of data: data can be accepted, but conclusions on the basis of data cannot. Howson waffles on this issue: "... we say nothing about whether it is correct to accept the data." Since any scientific data that I can imagine and take seriously (by which I mean to exclude such data as "I am now being appeared to redly," which I do not take to be a paradigm of 'scientific data') is corrigible, it seems to me that the same strictures should apply to 'exogenous given data' as apply to conclusions.

Leaving to one side the treatment of data, there seems to be no reason that one *can't* treat inductive inference in the way that Bayesians suggest. But not everyone agrees that this is the appropriate treatment for scientific or inductive or (for that matter) *practical* inference. There are a variety of formalisms that are being explored in philosophy and in computer science that are designed exactly to provide a way of arriving at conclusions that are to be regarded as corrigible. The various species of non-monotonic logic, default logic, logics of defeasible reasoning, and probabilistic inference in my sense, in which high probability warrants acceptance, are all logics designed to characterize what, in classical terms, must be regarded as 'invalid' inference. All of these approaches represent alternatives to the Bayesian approach.

That is exactly the problem, in the view of many Bayesians: Why should one endorse a method of inference that is *invalid*, which can lead from true premises to a false conclusion? No one would dream of endorsing a methodology incorporating principles of invalid inference in mathematics or in theology; why should one do so in science?

There are reasons. As Salmon argued in 1968, one reason is to provide the materials for the classical covering law view of explanation. To use a classical illustration, suppose the explanation for my broken car radiator is that there was no antifreeze in the water, that the temperature went down to 20 degrees last night, and that water expands on freezing, causing stresses that cannot be contained by automobile radiators. The covering laws involved here are that water freezes at 20 degrees, and that water expands on freezing. If we cannot accept these generalizations, we cannot accept the explanation.

There is no way to substitute a degree of belief for acceptance, and still have a covering law model of explanation. I may have a high degree of belief that water freezes at 20 degrees, but from this nothing follows about the water in my radiator. It is perfectly consistent with this belief that the water in my radiator does not freeze at 20 degrees.

Of course we can reject the covering law model of explanation, and replace it by Bayesian explanation: I have high degrees of belief about the propositions comprising the story, including a high degree of belief in its conclusion, that my radiator broke. We might be able to show that the assumption of high degrees of belief in the premises of the story entailed a high degree of belief in its conclusion. But this does not conform to the usual view of explanation.

Another difficulty with eschewing any form of inductive (i.e., risky) acceptance would be to explain engineering handbooks, which are far from being compendia of assertions about degrees of belief. Another is to do justice to the ordinary scientist who does not at all regard everything that he regards as *corrigible* as 'merely probable.' Thus if I am doing a computation that involves the mass of a proton, I'll look up the value in the latest handbook, and use the interval I am given there as if the mass were certain to fall in that interval. At the same time, I will not be shocked or dismayed if the next edition of the handbook contains a different value.

I take a natural and realistic view of science to allow for the acceptance of corrigible statements, both in the form of data and in the form of laws and hypotheses. Indeed, this is such a natural view that it is hard to see what motivates the Bayesian who wants to replace the fabric of science, already complicated enough, with a vastly more complicated representation in which each statement of science is accompanied by its probability, for each of us. Worse, all but a finite number of the empirical statements of our scientific language, as we have seen, must bear probabilities of 0 or 1, and thus cannot be corrected by the only procedure alleged to be warranted, Bayes' theorem. This appears to be a denial of even Bayesian corrigibility.

The reason, I think, that Bayesians have talked themselves into this odd position is that, like Hume, they seek a guarantee of correctness. They reject invalid forms of inference, and thus forms of inference that are inductive or nonmonotonic—that go beyond their premises in content. Probabilities, of course, are safe: no future experience can contravene a (subjective) probability statement. We cannot be mistaken about probabilities so long as they are subjective. We run no risk at all of being shown in error.

The history of inductive logic is in large part the history of attempts to convert induction to deduction. From Mill's methods on, inductive argument has sought validity. Russell (1948), Keynes (1921), and others have offered "postulates" which function to support inductive argument in the sense that they convert it to deduction. (Of course the postulate need not be deterministic; it can be phrased in terms of frequencies, and lead to conferring frequency or chance probability on inductive conclusions. Such argument is no less deductive than one employing deterministic a priori principles.) Arthur Burks and others have supported a view of induction according to which it rests

on "presuppositions"—a presupposition being nothing more or less than a bare faced assumption that allows us to convert an inductive argument into a deductive one.[5]

In the 1950's there was a general attempt to duck the problems of induction by talking instead about material 'rules of inference.' This may indeed bear the closest relation to the Bayesian proposal. According to this view (endorsed in various forms by Stephen Toulmin (1953, 1961), Gilbert Ryle (1937,1957), Peter Strawson (1952, 1959), and others), scientific reasoning is justified if it conforms to the norms for scientific inference. These norms embody various rules for making inferences that are material in the sense that whether the conclusions to which they lead are true or not depend on more than the truth of the premises—it depends on the nature of the world. For example, from the examination of a large and varied sample of crows, all of whom are black, we infer that all crows are black. We do *not* make use of any postulate to the effect that if we find that all the members of a large and varied sample of a population have property P, then all members of the population do. Such a postulate would require defense (and anyway, would be false). We just follow the rule.

This revolutionary approach to logic won few converts from philosophy in the long run. It was severely criticized (for example, by Cooley (1959)), on the grounds that replacing the conditional in the argument:

$$\text{If } P \text{ then } Q,$$
$$\underline{P}$$
$$\therefore Q$$

by a rule of inference, 'From P, Q may be inferred,' to obtain

$$\underline{P}$$
$$\therefore Q$$

doesn't really change the questions we can ask or the semantic justifications we can hope for.

None of these efforts to convert induction to deduction has succeeded. Nor could it. What we need is the analysis of the grounds on which we can reasonably leap beyond the data, not with any guarantee of success, nor even any guarantee of frequent success, but with confidence that our leap is *rationally defensible*. Postulates and presuppositions have been no help, for a variety of reasons, but not least for the reason that one man's presupposition is another man's fairy tale. Material rules of inference are no help, for if they are subject to criticism, they must be defensible, and if they are not, they are no better than presuppositions. If you and I disagree about the strength of a girder, I will not be convinced by being told your presuppositions.

7. Conclusion

Bayesianism in science is yet another effort to convert induction to deduction—to get plausible sounding conclusions that cannot be impugned by future events: to achieve *validity* for scientific inference. If I have a high degree of belief in h, relative to the evidence e, *that* conclusion is not impugned by the fact that relative to e and additional evidence e', I have a low degree of belief in h. That *is* the Bayesian conclusion: 'a posterior distribution [of belief]'. Bayesianism achieves validity at the cost of content. Fisher (quoted disapprovingly by Howson, p. 56) gave his view of subjective probability "... as measuring merely psychological tendencies, theorems respecting which are useless for scientific purposes." I think Fisher is perfectly correct, barring their hypothetical use for the purposes of psychology.

Bayesianism, as an general approach to scientific reasoning, must join the shattered hulks of all those previous failed attempts to make of inductive inference a

species of deduction. If scientific inference does not reach beyond what is minimally entailed by what happens to our sense organs, it is not worth our effort, and not worth our respect. We want to know what evidence is acceptable, and what suspect; we want to know what principles we can confidently employ in constructing better mousetraps, and what size the girders must be in our skyscrapers. We want to know what *is* the case, not what someone *believes* to be the case.

Notes

[1]Acknowledgment for support of research is due to the National Science Foundation.

[2]The distinction between "direct inference" and "inverse inference" is an old one; it was certainly well established in the 1920s when Fisher and Neyman were writing on the foundations of statistics.

[3]The charge of inconsistency is not difficult to dispose of: it depends on a confusion of use and mention, or quantifying into a referentially opaque context. It is a sentence *mentioning* the ratio r that occurs in the scope of the probability operator, and the ratio r itself that is the value of the probability expression.

[4]Even if the description is "the event at time *t* and place *p*," we know a lot, since we know of many things at places related to *p* and times related to *t*.

[5]The advantage of presuppositions is that they need not be defended: they are not *intended* to be defensible. More often, now, we hear that we can get nowhere without making assumptions, so what is important is to make the assumptions explicit. Again, this can be seen as a repudiation of responsibility: if I state something as an 'assumption' then I am not under an obligation to defend it.

References

Carnap, R. (1971), "A Basic System of Inductive Logic, Part I", in *Studies in Inductive Logic and Probability I*, Carnap and Jeffrey (eds.). Berkeley: University of California Press, pp. 33-165.

_____. (1950), *The Logical Foundations of Probability*. Chicago: University of Chicago Press.

Cooley, J.C. (1959), "Toulmin's Revolution In Logic," *Journal of Philosophy* 56: 297-319.

de Finetti, B. (1937), "La Prevision: Ses Lois Logiques, Ses Sources Sujectives", *Annales De L'Institute Henri Poincare* 7.

Fisher, R.A. (1956), *Statistical Methods and Scientific Inference*. New York: Hafner Publishing Co.

_____. (1930), "Inverse Probability", *Proceedings of the Cambridge Philosophical Society* 26: 528-535.

_ _ _ _ _ _ . (1924), "On a Distribution Yielding the Error Functions of Several Well Known Statistics", *Proceedings of the International Mathematical Mathematical Congress,* Toronto. pp. 805-812.

_ _ _ _ _ _ . *The Design of Experiments,* Hafner, New York, 1971. Probability Statistics Statistics First Edition 1935

Harman, G. (1986), *Change in View.* Cambridge: MIT Press.

Hintikka, J. (1966), "A Two-Dimensional Continuum of Inductive Methods", in *Aspects of Inductive Logic,* Hintikka and Suppes (Eds). Amsterdam: North Holland, pp. 113-132.

Howson, C. and Urbach, P. (1989), *Scientific Reasoning: the Bayesian Approach.* LaSalle, Ill: Open Court.

Jaynes, E.T. (1959), *Probability Theory in Science and Engineering; Colloquium Lectures in Pure and Applied Science 4, 1958* : Dallas, Texas, Socony Mobile Oil Corp. pp. 152-187.

Kyburg, H. and Pittarelli, M. (1992), "Some Problems for Convex Bayesians," *UAI–92,* Proceedings, pp. 149-154.

Levi, I. (1974), "On Indeterminate Probabilities", *Journal of Philosophy* 71: 39-418.

Lewis, D.K. (1980) "A Subjectivist's Guide To Objective Chance", *Studies in Inductive Logic and Probability II* , Jeffrey (Ed). Berkeley and Los Angeles: University of California Press, pp. 263-293.

Miller, D. (1966), "A Paradox of Information," *British Journal for the Philosophy of Science* 17: 59-61.

Russell, B. (1948), *Human Knowledge, Its Scope and Limits* . New York: Simon and Schuster.

Ryle, G. (1937), "Induction and Hypothesis", *Proceedings of the Aristotelian Society, Supplementary Volume* 16: 36-62.

_ _ _ _ . (1957), "Predicting and Inferring", in *The Colston Papers* 9, Korner (ed), pp. 165-170.

Savage, L.J. (1962),"Subjective Probability and Statistical Practise", *Foundations of Statistical Inference,* Barnard and Cox (Eds). New York: John Wiley and Sons, pp. 9-35.

Strawson, Peter F. (1952), *Introduction to Logical Theory,* London and New York: Methuen and Co.

Strawson, P.F. (1958), "On Justifying Induction". *Philosophical Studies* 9: . 20-21.

Toulmin, S. (1961), *Foresight and Understanding.* Bloomington: Indiana University Press.

_ _ _ _ _ _. (1953), *Philosopy of Science.* London: Hutchinson's University Library.