

## **DEBATES ON BAYESIANISM AND THE THEORY OF BAYESIAN NETWORKS**

**by Donald Gillies, King's College London\***

(Forthcoming in *Theoria*)

### **Contents**

- 1. The Origin, Development, and Establishment of the Paradigm of Classical Statistics (1900-1938)**
- 2. Fisher's and Neyman's Objections to Bayesianism**
- 3. Subjective Bayesianism overcomes Fisher's and Neyman's Objections: the Present State of the Problem**
- 4. Bayesian Networks and the Methodology of Testing**

Debates about the validity of Bayesianism have been a prominent feature of work on the foundations of statistics and on scientific method in general for most of this century. Nor do these debates show any sign of being resolved or discontinued, as is well illustrated by three recent books. Howson and Urbach's 1989 book argues that Bayesianism provides a sound foundation both for statistics and for scientific reasoning in general. By contrast Miller's 1994 book has this to say (125): 'Surely every one of the 46,656 varieties of Bayesianism catalogued by Good ... is vulnerable to at least one of the deadly objections raised in recent years.'; while Earman's 1992 book presents arguments for and against Bayesianism and develops an overall position which has some Bayesian features without being wholly Bayesian. In this paper I want to concentrate on

the Bayesian debate within statistics, although I will make use of general ideas drawn from philosophy of science. My main thesis is that there has been a significant change in the debate owing to the rise of subjective Bayesianism and recent results such as the theory of Bayesian networks. These developments have overcome the objections made to Bayesianism by the principal architects of classical statistics (Fisher and Neyman). Yet Bayesianism, I will argue, is still open to objections of a rather different kind, namely objections of a methodological character. I will illustrate these objections by considering some examples of Bayesian networks.

Section 1 of the paper sets the scene with a brief account of the origin, development, and establishment of the paradigm of classical statistics. This paradigm still dominates statistics, and one of its features is a rejection of Bayesian methods. Indeed the principal figures in the development of classical statistics (Fisher and Neyman) both criticized Bayesianism, although on different grounds. I will give an account of their objections to Bayesianism in section 2. It is important to realise that these objections were made against the classical/logical version of Bayesianism, a type of Bayesianism which has had a long history beginning with Bayes, and continuing through Laplace to Jeffreys, Keynes, and Carnap in the twentieth century. In the 1930's, however, a new type of Bayesianism appeared, the subjective Bayesianism of Ramsey and De Finetti. In section 3, I will try to show that this new version of Bayesianism overcomes the objections of Fisher and Neyman. This, I think, explains why there has been a new flowering of Bayesianism from the 1950's to the present. This flowering has resulted in many interesting developments, including the recent theory of Bayesian networks, a theory which has had numerous successful applications in the field of artificial intelligence. It is worth noting that Howson and Urbach (1989) adopt the position of subjective Bayesianism, and from this point of view criticize classical statistics. Does all this mean that we will soon have a change of paradigm, and that subjective Bayesianism will supersede the

methods of Fisher and Neyman as the dominant approach to statistics? It is of course always difficult to predict the future, but my own guess is that such a change of paradigm will not occur, for, although subjective Bayesianism has indeed overcome the objections made originally to Bayesianism, it is still liable to further objections of a rather different character. These might be characterised as *methodological* objections. I will conclude section 3 by giving an account of these objections, and then in section 4 I will examine their implications for the new theory of Bayesian networks. This theory was born in the paradigm of subjective Bayesianism, but I will maintain that it is possible to reformulate it in the framework of objective probabilities and the testing methodology of Fisher and Popper. I will try to show that this reformulation definitely improves the theory, so that, although the theory was created by the opposition group, it actually strengthens the established paradigm.

### **1. The Origin, Development, and Establishment of the Paradigm of Classical Statistics (1900-1938)**

I have selected 1900 as the starting point because it was in a paper published that year in which Karl Pearson introduced the chi-square test. I regard statistical tests as the core of classical statistics, so that the introduction of the first important statistical test is the crucial advance. In his 1908 paper, W.S.Gosset, who modestly wrote under the name 'Student', introduced the t-test. It was at this point that Fisher began his work. He gave a better mathematical foundation to the tests of Karl Pearson and Student, and introduced his own F-test, and the analysis of variance. In the 1920's he also developed the theory of statistical estimation. Two of Fisher's books were important for the introduction of the new ideas and techniques to statisticians. The first was his 1925 *Statistical Methods for Research Workers*, and the second his 1935 *The Design of Experiments*. In a paper of 1934 Neyman introduced another crucial

ingredient of classical statistics - his theory of confidence intervals. I have chosen 1938 as the final date because another important book was published that year - Neyman's *Lectures and Conferences in Mathematical Statistics*, in which he gave a general account of his position. Naturally there have been many discoveries and developments in classical statistics since 1938, but, by that date the paradigm had been established.

I have used the word 'paradigm' with deliberate reference to the philosophy of Kuhn, but this may seem a little strange. Statistics, after all, is a branch of methodology or scientific method concerned with how to reason about empirical data. Kuhn's theory, on the other hand, was designed to apply to branches of science concerned with developing theories about the natural world. Despite this difference in subject matter, I do think that many of Kuhn's ideas apply quite well to the development of statistics. To begin with, classical statistics has dominated since the 1940's, and still dominates today, in the sense that it is the approach used by a very large majority of statisticians in most of their work. It thus seems reasonable to say that the dominant paradigm<sup>1</sup> in the subject is classical statistics. There are moreover several other ideas of Kuhn's which can usefully be applied here. Kuhn emphasizes the importance of textbooks for teaching the paradigm, or, perhaps, as a cynic might say, for indoctrinating the new generation in the paradigm (Kuhn, 1962, 10). According to this *textbook criterion*, the content of a paradigm is more or less what appears in the standard textbooks. A most important book for classical statistics was Harald Cramér's *Mathematical Methods of Statistics*, published in Sweden in 1945, and in the United States in 1946. This is a most admirable book. The exposition is extremely clear, and contains many historical notes. It also made a very important mathematical improvement. Fisher used techniques of n-dimensional geometry in his original proofs. Cramér instead adopted the measure theory approach to probability. As a result his proofs were shorter and more comprehensible to the mathematical community. Granted then that

Cramér's book played an important rôle in establishing classical statistics as the dominant paradigm in the subject, it is interesting to examine what he has to say about Bayesianism. In fact only two of the 575 pages of his book make any reference to Bayesianism at all. Moreover his very brief remarks on the subject occur in a chapter entitled: 'Confidence Regions'. Cramér writes (1946, 507):

'In the older literature of the subject, probability statements of this type were freely deduced by means of the famous *theorem of Bayes*, one of the typical problems treated in this way being the classical problem of *inverse probability* .... However, these applications of Bayes' theorem have often been severely criticized, and there has appeared a growing tendency to avoid this kind of argument, and to reconsider the question from entirely new points of view. The attempts so far made in this direction have grouped themselves along two main lines of development, connected with the theory of *fiducial probabilities*, due to R.A.Fisher ... and the theory of *confidence intervals* due to J.Neyman ... . We shall here in the main have to restrict ourselves to a brief account of the latter theory.'

This passage illustrates another point of Kuhn's. In the creation of a paradigm, not all the ideas of the original researchers are included, but only that subset which is accepted by the majority of experts in the field. Beginning with his 1930 paper, Fisher published a series of works on the development of his theory of the *fiducial argument*. But this theory was not accepted by the general community of statisticians, who adopted instead Neyman's rival theory of *confidence intervals*. Fisher was very angry about this, but, everything considered, the community was right since the theory of confidence intervals does work, while the fiducial argument does not. Cramér who had studied with Fisher at Cambridge approaches the question in the most tactful way possible, saying in effect that limits of space do not allow a discussion of the fiducial argument, but this was, in effect, an exclusion of that theory from the paradigm.

Returning to the problem of Bayesianism, it is clear that Cramér's attitude is that of excluding Bayesianism completely from the paradigm. According to him, in cases in which in the past Bayesianism was employed, statisticians will in the future use the theory of confidence intervals. In effect Cramér rejects Bayesianism without however giving any criticism of this theory, but criticisms of Bayesianism were made by Fisher and Neyman. We shall consider these criticisms in the next section.

## 2. Fisher's and Neyman's Objections to Bayesianism

To explain these objections, let us take the most simple example. Suppose we are tossing a possibly biased coin, that the tosses are independent, and that the probability of heads is  $p$ , where  $p$  has some constant but unknown value in the interval  $[0, 1]$ . In the classical Bayesian approach, we begin with the assumption that we know *a priori* that  $p$  has a uniform distribution in the interval  $[0, 1]$ . This assumption is justified in the classical Bayesian approach by using the so-called *Principle of Indifference*. Since we have no reason to suppose that  $p$  is in one part of the interval rather than another, we should assign a uniform distribution to  $p$ .

Fisher, however, denies that we can have *a priori* knowledge, based on the Principle of Indifference, that  $p$  has a uniform distribution. His argument is as follows (see Fisher, 1956, 16-17). Define  $\phi$  by

$$p = \sin^2\phi \quad \text{where } 0 \leq \phi \leq \pi/2.$$

We have no reason to suppose that  $\phi$  is in one part of the interval rather than another. Thus, if the above argument based on the Principle of Indifference were

correct, we could equally well conclude that  $\phi$  had a uniform *a priori* distribution in  $[0, \pi/2]$ . But this would give to  $p$  a non-uniform *a priori* distribution in  $[0, 1]$ . Fisher therefore thinks that we cannot, in general, assign *a priori* distributions to parameters, and that Bayesianism fails for this reason to give a satisfactory account of statistical inference. Of course the point stressed here by Fisher is really a traditional one. It goes back to the paradoxes of geometrical probability, and the whole matter is well discussed in Keynes' *Treatise on Probability* (1921, 41-64).

The next argument against Bayesianism was, to the best of my knowledge, genuinely invented by Neyman, and appeared in his 1937. Fisher questioned whether we could know what *a priori* distribution a parameter had. Neyman went further and questioned whether it was admissible to ascribe an *a priori* distribution to a parameter at all.

To see the force of his argument, let us begin by observing that probability distributions can only be correctly ascribed to random variables. Now what can we regard as a random variable? The answer depends on what interpretation of probability we adopt. Let us begin by taking an objective interpretation defined as follows. Probabilities are taken to be objective features of the material world like charges or masses.<sup>2</sup> On this approach, we need to have a set of repeatable conditions, say  $\mathbf{S}$ . If any possible result of these conditions is given by a real number, we can say that the results are specified by a random variable  $X$ . The range of  $X$  is the set of possible results of  $\mathbf{S}$ , and the probability distribution of  $X$  gives the weight we attach to different possible results of  $\mathbf{S}$ . Probability distributions are related to the *objective* variation in the value of  $X$  obtained when  $\mathbf{S}$  is repeated.

Now can a parameter, such as  $p$  in the present example, be regarded as the result of repeating an underlying set of conditions? Sometimes perhaps it could

be. For example, suppose we have an urn filled with coins having different biases in different directions. The urn is shaken well, and a coin drawn from it. In this case  $p$  (the probability of heads for that coin) is the result of a set of repeatable conditions  $S$ . In general, however, when we are presented with a coin, its selection would not have been the result of a repeatable process of the sort just described. Our parameter  $p$  will not then be a quantity which fluctuates objectively on repeating some set of conditions. It will instead (if the background assumptions are correct) be a fixed, though unknown, constant. Thus, given an objective interpretation of probability of the kind described earlier, we cannot regard a parameter as a random variable and so cannot ascribe an *a priori* distribution to it as the Bayesian method requires. This is Neyman's new argument against Bayesianism.

For the sake of precision, we must make one small qualification to the above formulation of Neyman's argument. An unknown constant can, even given the above objective interpretation of probability, be regarded as a random variable in a 'trivial' or 'degenerate' sense. Suppose  $p$  has the value  $a$ , where  $a$  is a real constant. We can then say that  $p$  is a random variable with the distribution

$$\begin{aligned}\text{Prob}(p = a) &= 1 \\ \text{Prob}(p \neq a) &= 0.\end{aligned}$$

In his 1937 paper, Neyman rules out such degenerate random variables by requiring that the range of any random variable should contain two subsets  $A$  and  $B$  such that

$$A \cap B = \emptyset, \text{Prob}(A) \neq 0, \text{Prob}(B) \neq 0 \quad (1)$$

However, it seems to me mathematically convenient to be able to treat a constant as a random variable by the above method. Let us therefore allow such random

variables but say that a random variable is trivial or degenerate unless Neyman's condition (1) is satisfied. We can then put Neyman's point by saying that parameters will in general be unknown constants and hence only random variables in a degenerate sense. They will not be random variables in the non-trivial sense which is required by Bayesian inference.

In formulating this argument, Neyman was, I think, influenced by von Mises whose 1928 he helped to translate from German for the first English edition of 1939. Von Mises bases his approach to probability on the concept of *collective* which he describes as a long (1928, 12) 'sequence of uniform events or processes which differ by certain observable attributes, say colours, numbers, or anything else.' Von Mises argues that probabilities in the mathematical sense can only be defined within collectives. He says (1928, 12):

'The principle which underlies the whole of our treatment of the probability problem is that a collective must exist before we begin to speak of probability.'

Actually this formulation is perhaps a little too strong, because von Mises does not deny that we sometimes speak of probabilities in ordinary language where there is no collective. His thesis is that such probabilities cannot be brought within the scope of the mathematical theory. This view he illustrates by an analogy with the concept of work in mechanics. Work in the precise mathematical sense is defined as force times distance, and when work is used in the mathematical theory it is always used in this sense. This excludes some every day uses of work. For example if a waiter holds up completely steadily for some time a heavy tray laden with canapes, we could quite reasonably say that he has done a lot of hard work, but he has not done any work in the mathematical sense used in mechanics. Similarly, von Mises argues, the probabilities to which the mathematical theory applies must all be defined in a collective, even though

the word probability might sometimes be used in a wider sense in ordinary language. As he puts it (1928, 15):

“The probability of winning a battle’, for instance, has no place in our theory of probability, because we cannot think of a collective to which it belongs. The theory of probability cannot be applied to this problem any more than the physical concept of work can be applied to the calculation of the ‘work’ done by an actor in reciting his part in a play.’

Using von Mises’ framework, Neyman’s objection to Bayesianism can be put like this. An unknown parameter such as  $p$  is not in general a member of any collective. So we cannot assign to it a probability distribution and apply the mathematical theory of probability.

Neyman’s argument requires stronger premisses than Fisher’s, but it also rules out more. In particular, if we accept Neyman’s argument, we should not assign any probability distribution, whether *a priori* or *a posteriori*, to an unknown parameter. So the argument rules out Fisher’s fiducial argument as well as Bayesianism, but it does allow confidence intervals. We see then that the argument underlies the debate between Fisher and Neyman concerning the respective merits of the fiducial argument and confidence intervals.

### **3. Subjective Bayesianism overcomes Fisher’s and Neyman’s Objections: the Present State of the Problem**

As I have already remarked, the objections of Fisher and Neyman were made to the classical/logical version of Bayesianism which has had a long history beginning with Bayes, and continuing through Laplace to Jeffreys, Keynes, and Carnap in the twentieth century. In the early 1930’s, however, just

at the time when the paradigm of classical statistics was being consolidated, the first papers developing a new approach to Bayesianism (the subjective approach) were published. This is an example of a simultaneous discovery, because it was made by Ramsey at Cambridge in England, and by De Finetti in Italy. Ramsey read a paper on the subject at Cambridge in 1926, and it was published, after his premature death, in the 1931 collection of his papers. De Finetti wrote an account of the subjective approach to probability in 1928, and published his first paper on the question in 1930. A more extensive account of his views is contained in his (1931a & b), published in the same year as Ramsey's posthumous collection. The two discoverers worked quite independently of each other. Ramsey almost certainly never heard of De Finetti, and De Finetti only read Ramsey's paper on subjective probability several years after his own views had been formulated and published. In fact there are quite significant differences between Ramsey's approach to the question and De Finetti's. These are discussed in an interesting article by Galavotti (1991), and Sahlin's 1990 book on Ramsey's philosophy contains some valuable observations on this question (41-2 & 53). However, for the purposes of this paper I can concentrate on points which are common to both Ramsey and De Finetti.

The subjective theory interpreted probability as the degree of belief of a particular individual (Mr R say). The first important step in its development was to introduce a method for measuring this degree of belief. The measure proposed was the rate at which Mr R would bet under certain conditions. The second important step was to show, using the very reasonable condition of coherence, that these betting rates (or quotients) must obey the mathematical axioms of the probability calculus. This result is the famous Ramsey-De Finetti theorem.

Now when von Mises suggested in 1928 that the mathematical theory of probability should be limited to probabilities defined within collectives, his

position was a very reasonable one. At that time, there was no satisfactory way of extending the mathematical calculus to probabilities in any more general sense. The development of the subjective theory, and, in particular, the Ramsey-De Finetti theorem changed the situation. It now became possible to introduce probabilities as subjective betting quotients, even if no collective was defined, and to show that these betting quotients must obey the standard mathematical axioms of probability. We can illustrate this by taking von Mises' own example of the probability of winning a battle. Suppose a battle is due to be fought tomorrow. Von Mises is right to say that it would be very difficult to define an appropriate collective, because battles are individual events, fought under widely different conditions. However Mr R can certainly bet on the outcome of the battle, and so we can introduce his subjective probability of a particular side winning.

Subjective probabilities overcome Neyman's objection, which, as I argued, was based on von Mises' ideas. An unknown parameter  $p$  may indeed not be the result of a set of conditions whose repetition produces objective random variation. Thus we may not be able to assign a probability distribution to the parameter if the probabilities are understood in an objective, scientific sense. However, we can certainly introduce the subjective probability of Mr R that  $p$  will lie in a particular interval  $[a,b]$  say. This is simply the rate at which Mr R will agree to bet that  $p$  lies in  $[a,b]$ . It might be objected that we can never know for certain what the value of an unknown parameter is, and so the bet as just described can never be settled. If, however, a sufficiently large sample is taken, this will fix the value of  $p$  within agreed limits which are sufficiently narrow to allow a decision to be reached as to whether  $p$  lies in  $[a,b]$  or not. Thus the bet could be settled at least in principle, and so can be regarded as legitimate. By considering bets for all the relevant intervals, we can introduce Mr R's *a priori* subjective probability distribution for  $p$ . So Neyman's objection to assigning probability distributions to unknown parameters is overcome,

provided the probability distributions assigned are understood in the subjective sense.

Fisher's objection can also be overcome in the subjective approach. In the older classical/logical tradition, it was necessary to use the Principle of Indifference to obtain an *a priori* probability distribution. Moreover, since such a distribution was supposed to have a logical character, it had to be unique. So when different *a priori* distributions were obtained using the Principle of Indifference (as in Fisher's example), this amounted to a contradiction. All this changes in the subjective approach. Here there is no longer any need to use the Principle of Indifference to obtain an *a priori* probability distribution, for such a distribution simply represents the opinion of a particular individual. It is possible in theory at least for an individual to choose any distribution subject only to the constraint of obeying the probability axioms obtained from the condition of coherence. There is no contradiction in Mr R choosing one *a priori* distribution, and Mr D-F another. They simply have different opinions on the question, and there is no unique correct *a priori* distribution obtainable by some kind of logical argument. It is for these reasons that Ramsey writes (1931, 189):

‘ ... the Principle of Indifference can now be altogether dispensed with; ... To be able to turn the Principle of Indifference out of formal logic is a great advantage; for it is fairly clearly impossible to lay down purely logical conditions for its validity, as is attempted by Mr Keynes.’

These successes of the subjective conception of probability mean, in my opinion, that there has been what Lakatos would have called a problem shift in the Bayesian debate. The old debate concerned the question of *a priori* distributions. Was it possible to know what these distributions were? (Fisher) Was it legitimate at all to introduce such distributions for unknown parameters? (Neyman) Subjective Bayesianism found quite convincing replies to these

difficult questions. But this does not mean that all the differences between Bayesianism and Classical Statistics have been resolved: it means only that other points and other differences are now more important than the question of *a priori* distributions. What then is the present state of the problem?

I think there are two important differences between Classical Statistics and subjective Bayesianism: (i) a difference in the interpretation of probability - objective in one case, and subjective in the other, and (ii) in close connection with (i) a difference in methodology. This methodological difference seems to me the more important, and indeed I regard it as the crucial point in the current state of the Bayesian debate.

Let us begin with Classical Statistics. This adopts an objective conception of probability. Probabilities are regarded as aspects of the material world similar to the masses of bodies. A probability or a mass has an objective value which might be unknown, but which exists nonetheless. If the distribution of the values of a random variable is unknown, we can make a conjecture as to what it is, but such conjectures must be subjected to statistical tests, and it could well happen that these tests show that the conjecture is incorrect. In effect, as I have already claimed, the key part of Classical Statistics is the theory of statistical tests. Moreover the methodology of Classical Statistics is precisely Popper's method of conjectures and refutations.

If we now turn to subjective Bayesianism, we have a very different picture. Probabilities are the opinions of a particular individual, for example of our Mr R. If Mr R judges at a particular moment  $t$  that the probability of a particular event  $A$  is, let us say,  $3/4$ , it makes no sense to say afterwards that this value has been shown to be wrong by experience. Whatever happens, it remains true that Mr R at  $t$  had a degree of belief of  $3/4$  in  $A$ . If new evidence  $E$  comes to light between  $t$  and the present, Mr R will change his *a priori* probability  $P(A)$

to an *a posteriori* probability  $P(A | E)$ , but without repudiating his original opinion. The change from  $P(A)$  to  $P(A | E)$  is made using Bayes' Theorem, and is called *Bayesian conditionalisation* (or *conditioning*). So the key difference between the two approaches is a methodological difference - statistical tests (conjectures and refutations) versus Bayesian conditionalisation. De Finetti describes this difference with great clarity. The passage is the following (De Finetti, 1937, 146. I have slightly altered De Finetti's notation to agree with that used in the present paper):

'Whatever be the influence of observation on predictions of the future, it never implies and never signifies that we *correct* the primitive evaluation of the probability  $P(A)$  after it has been *disproved* by experience and substitute for it another  $P^*(A)$  which *conforms* to that experience and is therefore probably *closer to the real probability*; on the contrary, it manifests itself solely in the sense that when experience teaches us the result  $E$  on the first  $n$  trials, our judgment will be expressed by the probability  $P(A)$  no longer, but by the probability  $P(A | E)$ , i.e. that which our initial opinion would already attribute to the event  $A$  considered as conditioned on the outcome  $E$ . Nothing of this initial opinion is repudiated or corrected; it is not the function  $P$  which has been modified (replaced by another  $P^*$ ), but rather the argument  $A$  which has been replaced by  $A | E$ , and this is just to remain faithful to our original opinion (as manifested in the choice of the function  $P$ ) and coherent in our judgement that our predictions vary when a change takes place in the known circumstances.'

I will now give my criticism of subjective Bayesianism. It is that subjective Bayesianism is too conservative. Once our Mr R has decided in favour of an initial distribution, that distribution is changed only by means of Bayesian conditioning; but there are many cases in which such changes are not sufficiently radical. Using Bayesian conditioning it is possible to change for example the values of parameters, but the basic framework of the model implied

by the original choice of distribution remains unchanged. So Bayesianism is all right for changes *within a fixed framework*, but it does not allow changes of the basic framework.

Quite to the contrary, the testing methodology of Fisher and Popper requires that every assumption of the underlying framework be checked by statistical tests. If one of these tests gives a negative result (a refutation), it becomes necessary to change some assumption, and introduce a new assumption, perhaps based on some new idea or conception. Often changes of this type lead to models which work better. Subjective Bayesianism blocks this type of progress.

This argument can be summarised as follows. Subjective Bayesianism only allows change by means of Bayesian conditioning.<sup>3</sup> This is too conservative, since such changes are often too small. Bayesian conditioning does not change the basic structure of the original opinion, but often this structure must be changed to allow progress. The testing methodology of Fisher and Popper forces us to criticize the basic assumptions, and to check them with statistical tests. The refutation of some of these assumptions often leads to better assumptions and models; that is to an important form of progress which cannot take place within the framework of subjective Bayesianism.

This is my general argument against subjective Bayesianism. In the final section of the paper, I would like to illustrate it with an example drawn from the new theory of Bayesian networks.

## 5. Bayesian Networks and the Methodology of Testing

Bayesian networks were introduced in the late 1980's and early 1990's [Pearl (1988), Spiegelhalter and Lauritzen (1990), and Neapolitan (1990)]. I believe that they represent a most important advance in probabilistic reasoning for artificial intelligence, and, in particular, for expert systems. At least two of the principal contributors (Pearl and Spiegelhalter) are subjective Bayesians, and there is thus no doubt that subjective Bayesianism has proved heuristically very fruitful. Despite this, however, the theory of Bayesian networks is not completely tied to the subjective Bayesian approach. It is possible to translate it into the Fisher-Popper framework. I will begin this section with an explanation of how this translation can be carried out, and I will then try to show with a practical example that the translation brings some advantages.

Let us consider the field of medicine. A doctor has a patient with a group S say of symptoms, and the doctor considers the probability that the patient has a disease D say. The doctor, who is also a mathematician, writes down Bayes' theorem.

$$P(D | S) = \frac{P(S | D) P(D)}{P(S)}$$

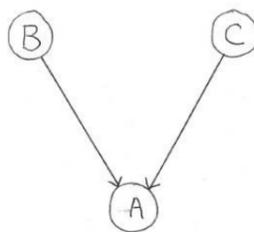
If the doctor is a subjective Bayesian, all these probabilities will represent his degrees of belief, so that, for example,  $P(D | S)$  will be the doctor's degree of belief that the patient has disease D given that the patient has symptoms S. This interpretation is not, however, necessary for it is possible to interpret all these probabilities as objective probabilities.  $P(D)$  is the probability of the disease in the population, which may well be known on the basis of statistical data.  $P(S | D)$  is the probability that a person who has the disease in question also has the group S of symptoms. This might not be known, but it could in principle be calculated from statistical data about people suffering from the disease. Finally

$P(S)$  is the probability that a person chosen at random from the population has the group of symptoms  $S$ . This is undoubtedly an objective probability. It would perhaps be difficult to calculate its value from statistical data, but there might be methods of forming an estimate of this value. It is clear then that all the probabilities which are used in the field of medicine can be interpreted as objective. The only difficulty would be to calculate the values of these probabilities from statistical data in some cases. However this problem is not an insuperable one.

I have just given a simple example in which it is possible to use Bayes' theorem. In more difficult cases, however, it becomes necessary to use a more complicated structure - a Bayesian network. The general definition of a Bayesian network is a little complicated (see Neapolitan, 1990, 158-9), but I can illustrate the point I want to make with a quite simple example, which I call: 'The Englishman in summer'. The example is taken from Neapolitan (1990, 186-7), but the name is my own invention! The network is illustrated in Figure 1.

Figure 1

A simple example of a Bayesian Network. The Englishman in summer.



A = Mr E sneezes

B = Mr E has a cold

C = Mr E has hay fever

Assumption of Independence: B and C are independent, i.e.  
 $P(B|C) = P(B)$  e  $P(C|B) = P(C)$ .

The interpretation of the network in the framework of subjective Bayesianism is the following. The medical expert observes the event A, and knows that it could be caused by B or by C. This is shown in the figure by the arrows which go from B to A, and from C to A. The doctor also knows that the possible causes B and C are independent. A cold does not cause hay fever or vice versa. This is shown in the figure by the fact that there is no connection between B and C. This causal independence is then used as a reason for supposing that there is probabilistic independence, i.e.  $P(B | C) = P(B)$  and  $P(C | B) = P(C)$ . The doctor then adds to the network his subjective *a priori* probabilities, giving, for example, his degree of belief in B given A [ $P(B | A)$ ]. These initial probabilities can be changed by means of probability calculations, when further facts become known. The theory of Bayesian networks explains how these calculations should be carried out, not just in simple cases like this, but in much more complicated cases as well.

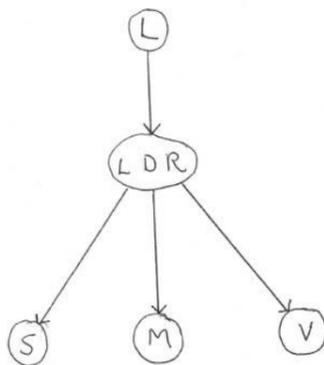
This simple example illustrates a fundamental point. Every Bayesian network involves assumptions of independence or of conditional independence. These assumptions are necessary. Without them the problem becomes computationally intractable. In the subjective Bayesian approach these assumptions are justified by the knowledge of the expert, and, in particular by his knowledge of causal chains. What would be the Popperian attitude to this procedure? A Popperian would not deny that it could be very useful to use the expert's knowledge, especially of causal chains, to construct the Bayesian network with its assumptions of independence or conditional independence. But for a Popperian this initial opinion is only a conjecture. This conjecture might be true or it might be false. It is therefore very important to subject the conjecture to statistical tests to see whether it is confirmed or refuted. In particular, it is essential to check the assumptions of independence or conditional independence by statistical tests. This is the key point. In the framework of subjective Bayesianism, the assumptions of independence and conditional independence in

a Bayesian network are not checked by tests.<sup>4</sup> I will now try to show with a practical example that it is useful to carry out such tests for two reasons. First it is possible that the assumptions of independence or conditional independence might be false, and, second, if they are false, it may be possible to substitute for them other assumptions which give a Bayesian network which is simpler and more efficient.

The example is taken from a research project in artificial intelligence in which I participated with Duncan Gillies (my brother) and Enrique Sucar in the Department of Computing, Imperial College, London (Sucar, Gillies, and Gillies, 1993). The research concerned a medical instrument called an 'endoscope'. This allows a doctor to put into the colon of a patient a small camera which transmits an image of the interior of the colon to a television screen. In this image an expert can recognise various things in the interior of the colon. Let us take two such things as examples. One is called the 'lumen' which is the opening of the colon. Despite its name, it generally appears as a large dark region; but sometimes it is smaller and surrounded by concentric rings. Another is called a 'diverticulum' and is a small malformation in the wall of the colon, which can cause some illnesses. A diverticulum generally appears as a dark region, smaller than the lumen, and often circular. It is a problem then to program a computer to recognise from the image the lumen or a diverticulum. This is a typical problem of computer vision. To solve it, we constructed a Bayesian network with the help of an expert in medical endoscopy. Figure 2 shows only a small part of this network, but it is sufficient to illustrate the points which I would like to make.

Figure 2

A part of the Bayesian Network for recognising the lumen



L = Lumen

LDR = Large Dark Region

S = Size

M = Mean of Illumination

V = Variance of Illumination

Assumptions of Conditional Independence: S, M, V are conditionally independent of L and mutually conditionally independent, given LDR.

In every Bayesian network certain assumptions of independence or conditional independence are by definition satisfied. In this case, S, M, V must be conditionally independent of L and mutually conditionally independent, given LDR. Using the testing methodology of Fisher-Popper, we considered these assumptions as conjectures which needed to be checked by statistical tests. In fact it was not difficult to carry out a statistical test. We took a random sample of 300 images of the colon. Entire sessions investigating the colon with the endoscope are recorded on video as a standard practice; so there is no shortage of data for these problems. In every image of the colon, the lumen, if it was visible, was identified by an expert. It was also possible to calculate LDR, S, M, V and so to form estimates of the various probabilities in the network using observed frequencies. To carry out a statistical test of the assumptions of conditional independence, we calculated the correlation between S, M, V

conditional on LDR. We used two correlation coefficients, Pearson's  $r$ , and Kendall's  $\tau$ . The results are given in Table 1.

Table 1

**Correlation between the parameters S, M, V conditional on LDR**

Correlation Coefficients	S & M	S & V	M & V
Pearson's $r$	-0.146	0.264	0.482
Kendall's $\tau$	-0.089	0.116	0.342

It can be seen that the correlations between S & M, and S & V are fairly small. So we can regard the assumptions of conditional independence between S & M, and S & V as having passed this statistical test, and we can continue to adopt them. The case of M & V is very different. Here the correlations are rather large. So the assumption of conditional independence has failed the statistical test and should no longer be adopted.

But what should be done in this situation? The simplest response was to eliminate one of the two parameters M and V on the grounds that, since they were correlated, only one could give almost as much information as both. The results of this elimination are given in Table 2, which was prepared using another random sample of more than 130 images of the colon.

Table 2

The Results of Eliminating one of the two Correlated Parameters

	Percentage of Correct Results	
	Lumen	non-Lumen
All parameters (S, M, V)	89%	54%
Elimination of V (S, M)	93%	79%
Elimination of M (S, V)	97%	50%

It can be seen that the elimination of one of the parameters gave better results than those obtained using all three parameters. At first sight this seems a paradox, because these better results were obtained using less information. The explanation is simple however. Undoubtedly there is more information in all three parameters (S, M, V) than in only two (for example S, M). But the greater amount of information in the three parameters was used with mathematical assumptions of conditional independence which were not correct. The lesser amount of information in the two parameters S, M was, by contrast, used with true mathematical assumptions. So less information in a correct model worked better than more information in a mistaken model. Moreover, since the modified Bayesian network was simpler, the calculations using it were carried out more quickly. So, to conclude, the modified Bayesian network was more efficient, and gave better results. This shows the value of using the testing methodology of Fisher-Popper.

Naturally this example is very simple, but it suggests an interesting research programme which could lead to useful results. This programme does

not completely contradict the theory of Bayesian networks developed in the framework of subjective Bayesianism, but rather suggests an addition which could improve the theory. In this approach the methods used now by the subjective Bayesians to construct Bayesian networks are considered as a first phase - the heuristic phase. A second phase is then added, that of testing and modification, in which the network is subjected to statistical tests, particularly of the assumptions of independence and conditional independence. If some of these assumptions are shown to be false, the network is modified by eliminating them, and adding better assumptions. But how should these tests and modifications be carried out? I have given a simple example, but a great deal of work would be necessary to discover how to carry out tests and make appropriate modifications in the various cases, many much more complicated, which are useful in practice. This work constitutes the research programme suggested by the application of the testing methodology of Fisher and Popper to Bayesian networks.

**NOTES**

- \* An earlier version of this paper was read at the Istituto per le Applicazioni del Calcolo “Mauro Piccone” (IAC) in Rome in October 1996. I found the lively discussion and comments made on that occasion very useful in developing the paper. Wlodek Rabinowicz and Jon Williamson were kind enough to read the paper and offer detailed comments, several of which have been incorporated into the final version. I would also like to acknowledge the help of a grant from the Humanities Research Board of the British Academy which provided me with some research leave during which this paper was written.
1. In his analysis of the natural sciences, Kuhn claims that, apart from exceptional revolutionary periods, only one paradigm exists at a particular time in a mature natural science. Whether or not this is true of mature natural sciences, it does not seem to me true of statistics. In this field there have been for the last fifty or so years, two paradigms. The paradigm of classical statistics has undoubtedly dominated, but the Bayesian paradigm has constituted a strong and intellectually stimulating opposition.
  2. In this paper when we speak of objective probabilities or an objective interpretation of probability, ‘objective’ will always be used in the sense here explained, i.e. as referring to something which exists in the material world. There is another sense of objective probability which appears in Keynes’s logical theory of probability (see his 1921). According to Keynes, there is, given some hypothesis  $h$  and a body of evidence  $e$ , a single correct degree of belief in  $h$  conditional on  $e$ . Probability is

interpreted as this rational degree of belief. Different individuals possessing the same evidence  $e$  may diverge from the rational degree of belief in  $h$ , and have different individual degrees of belief. Such degrees of belief are for Keynes subjective, while the rational degree of belief is objective (see 1921, 4). This is clearly a different sense of objective, because Keynes is not suggesting that rational degrees of belief exist in the world of nature. Keynes's sense of objective will not be used in this paper.

3. It might be objected that a version of subjective Bayesianism could be developed in which changes of belief by means other than Bayesian conditioning are allowed. This is certainly possible, though it has not yet been done in detail to the best of my knowledge, and there are difficulties in the way of carrying out the programme. It would have to be specified under what circumstances the switch is made from Bayesian conditioning to another form of belief change, and also what form or forms these other kinds of belief change would take.
4. It might be objected that a method of testing could be developed within the framework of subjective Bayesianism. As far as I know this has not been done as yet, though I would certainly welcome such a development which would bring subjective Bayesianism closer to the position advocated here.

## REFERENCES

- Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton University Press. 1961.
- De Finetti, B. (1930). Fondamenti Logici del Ragionamento Probabilistico, *Bollettino dell'Unione Matematica Italiana*, **5**, 1-3.
- De Finetti, B. (1931a). *Probabilism*. English translation in *Erkenntnis*, 1989, **31**, 169-223.
- De Finetti, B. (1931b). On the Subjective Meaning of Probability. English translation in De Finetti, 1993, *Induction and Probability*, Clueb-Bologna, 291-321.
- De Finetti, B. (1937). Foresight: Its Logical Laws, Its Subjective Sources. English translation in H.E.Kyburg and H.E.Smokler (eds.) *Studies in Subjective Probability*, John Wiley, 1964, 93-172.
- Earman, J. (1992). *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*. The MIT Press.
- Fisher, R.A. (1925). *Statistical Methods for Research Workers*. Oliver & Boyd.
- Fisher, R.A. (1930). Inverse Probability, *Proceedings of the Cambridge Philosophical Society*, **26**, 528-35.

- Fisher, R.A. (1935). *The Design of Experiments*. Oliver & Boyd.
- Fisher, R.A. (1956). *Statistical Methods and Scientific Inference*. 2nd Edition. Oliver & Boyd. 1959.
- Galavotti, M.C. (1991). The Notion of Subjective Probability in the Work of Ramsey and De Finetti, *Theoria*, **LVII**(3), 239-59.
- Howson, C. and Urbach, P. (1989). *Scientific Reasoning: The Bayesian Approach*. Open Court.
- Keynes, J.M. (1921). *A Treatise on Probability*. Macmillan. 1963.
- Kuhn, T.S. (1962). *The Structure of Scientific Revolutions*. University of Chicago Press. 1969.
- Miller, D.W. (1994). *Critical Rationalism. A Restatement and Defence*. Open Court.
- Neapolitan, R.E. (1990). *Probabilistic Reasoning in Expert Systems*. John Wiley.
- Neyman, J. (1934). On the Two Different Aspects of the Representative Method: the Method of Stratified Sampling and the Method of Purposive Selection, *Journal of the Royal Statistical Society*, **97**, 558-625.
- Neyman, J. (1937). Outline of a Theory of Statistical Estimation based on the Classical Theory of Probability. In *A Selection of Early Statistical Papers of J. Neyman*, 250-90, Cambridge University Press, 1967.

- Neyman, J. (1938). *Lectures and Conferences on Mathematical Statistics and Probability*. 2nd Edition. Washington. 1952.
- Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann.
- Pearson, K. (1900). On the Criterion that a given System of Deviations from the Probable in the case of a Correlated System of Variables is such that it can be reasonably be supposed to have arisen from Random Sampling. Reprinted in *Karl Pearson's Early Statistical Papers*, Cambridge University Press, 1956, 339-357.
- Ramsey, F.P. (1931). *The Foundations of Mathematics*. Routledge & Kegan Paul. 1965.
- Sahlin, N.-E. (1990). *The Philosophy of F.P.Ramsey*. Cambridge University Press.
- Spiegelhalter, D.J. and Lauritzen, S.L. (1990). Sequential Updating of Conditional Probabilities on Directed Graphical Structures, *Networks*, **20**(5), 579-605.
- Student (Gosset, W.S.) (1908). The probable error of a mean, *Biometrika*, **6**, 1-25.
- Sucar, L.E., Gillies, D.F. and Gillies, D.A. (1993). Objective Probabilities in Expert Systems, *Artificial Intelligence*, **61**, 187-203.

Von Mises, R. (1928). *Probability, Statistics and Truth*. 2nd revised English edition. Allen & Unwin. 1961.