CHAPTER 11

Objections to the Subjective Bayesian Theory

a INTRODUCTION

In the preceding chapters we have developed the theory of subjective or personalistic Bayesianism as a theory of inductive inference. We have shown that it offers a highly satisfactory explanation of standard methodological lore in the domains of both statistical and deterministic science; and we have also argued at length that all the alternative accounts of inductive inference—like Popper's or Fisher's—achieve their explanatory goals, where they achieve them at all, only at the cost of quite arbitrary stipulations. However, the subjective Bayesian theory itself has been the object of much critical attention, to such an extent, in fact, that it is still regarded in some influential quarters as vitiated by hopeless difficulties. These difficulties, in our view, stem from misunderstanding and confusion, and in this final chapter we shall do our best to dispel both.

Of the standard criticisms some—due largely to Popper and his followers—are answered relatively simply and quickly, and we shall deal with these first. We shall then consider an objection, due to Clark Glymour, which points to an apparently insuperable problem in explaining within any Bayesian theory how a hypothesis can be supported by data already known at the time the hypothesis was proposed. We often do want to say that hypotheses may be so supported: it is, after all, something of a commonplace that Einstein's General Theory of Relativity was supported by the value already accepted at the time of the seconds of arc through which Mercury's perihelion annually precesses; and the reader can no doubt think of many other examples. Glymour's objection, however, is precisely that the Bayesian theory is incapable of explaining how any data can support a theory proposed after the data became known.

We shall argue that Glymour's objection is false: the Bayesian theory can explain how data already known can support theories. However, the successful rebuttal of Glymour's objection appears to bring another in its train. This is that the Bayesian is incapable of discriminating, in his assessment of the support of a hypothesis by evidence, between evidence obtained independently of that hypothesis and evidence the hypothesis was deliberately constructed to explain. Then, runs the objection, and it seems prima facie a very powerful one, the Bayesian theory must be incorrect since quite clearly there should be no support of the hypothesis by the evidence in the second case. We shall show that despite its apparent plausibility, this last claim is incorrect, and that on the contrary there are many well-known examples of scientific theories drawing support from data which they were constructed to explain. Moreover, it turns out that the Bayesian approach reproduces exactly the sorts of informal reasoning actually employed in cases like these.

Though they come last, the remaining three objections have been the most influential. One, which has dominated the discussion for practically the whole of the present century, is that a subjective, degree-of-belief interpretation of the probabilities in Bayes's Theorem is inadequate precisely because it would make science a purely subjective affair. How then, it is objected, can the subjectivist explain the widespread agreement that science is correctly opposed to superstition in its claims on our credence because and only because it is based on objectively justifiable canons of inference, not on what people, perhaps whimsically, actually do believe and the extent to which they believe it? We shall argue that this objection rests on a confusion, and that a Bayesian reconstruction of the procedures of inductive inference poses no threat whatever to the objectivity of the scientific enterprise.

The next objection concerns the principle of conditionalisation, that is to say the principle that if $P(h \mid e)$ is your conditional probability of h on e, and you learn e (but nothing stronger), then consequent upon this information, your degree of belief in h is, if you are consistent, equal to $P(h \mid e)$. A frequently made charge is that the use of this principle commits the Bayesian to the unrealistic and certainly unwelcome existence of some *Ur*-distribution, from which his current distribution of belief is obtained by successive applications of the principle to incoming data. He is further committed, it is alleged, to the equally unrealistic supposition that all those items of data must be absolutely certain, since no allowance is made for conditioning on data which do not have this apodeictic character. We shall show that this multiple objection is a consequence of multiple confusions about the claims and aims of the subjective Bayesian approach, and that when these are dissipated, so too are the objections with them.

The final objection we shall discuss is that even were there nothing else wrong with this Bayesian theory, the empirically demonstrable fact is that people simply do not make their considered judgments according to its prescriptions. Were this true, and the evidence seems unequivocal, then it would appear to follow that the status of the subjective Bayesian theory as explanatory of the methodological evaluations people actually make must be severely in doubt. We shall show in due course that this conclusion is incorrect; let us now proceed to review all these objections in turn.

b THE BAYESIAN THEORY IS PREJUDICED IN FAVOUR OF WEAK HYPOTHESES

Discussing theories of inductive inference which assess the empirical support of hypotheses by changes in their probabilities on receipt of the relevant new data, Watkins (1987, p. 71) asserts that such theories are "prejudiced" in favour of logically weaker hypotheses. This is a favourite charge of the Popperian school and is frequently made by its eponymous founder; for example, Popper (1959, p. 363; his italics) writes that "[scientists] have to choose between high probability and high informative content, since *for logical reasons they cannot have both*".

Such a charge is quite baseless. There is *nothing* in logic or the probability calculus which precludes the assignment of even probability 1 to any statement, however strong, as long as it is not a contradiction. The only other way in which probabilities depend on logic is in their decreasing monotonically from entailed to entailing statements. But this again does not preclude anybody from assigning any consistent statement as large a probability as they wish. Popper's thesis that a necessary concomitant of logical strength is low probability is simply incorrect.

Glymour attempts to argue a variant of Popper's objection, but this too is easily repulsed. Glymour claims that since the observable consequences of scientific theories are at least as probable as the theories themselves, then in a Bayesian account one is unable to account for our entertaining theories at all:

> On the probabilist view, it seems, they are a gratuitous risk. The natural answer is that theories have some special function that their collection of observable consequences cannot serve; the function most frequently suggested is explanation.... [But] whatever explanatory power may be, we should certainly expect that goodness of explanation will go hand in hand with warrant for belief, yet if theories explain and their observational consequences do not, the Bayesian must deny the linkage. (Glymour, 1980, pp. 84–85)

The Bayesian certainly does want to justify the quest for theories in terms of a desire for explanation that a congeries of observational laws cannot by itself provide; but he would also, for very good reason, deny the linkage Glymour alleges between explanatory power and warrant for belief. Indeed, counterexamples to the claim that any such linkage exists are only too easy to find: a tautology, to take an obvious one, has maximal warrant for belief and minimal explanatory power. This does not, of course, imply that what we take to be good explanations do not tend to have correspondingly high probabilities on the available evidence. They do. But Glymour's premiss makes the additional claim that an increase in "warrant for belief" should imply an increase in explanatory power. That premiss is clearly false, and Glymour's objection collapses.

It is odd that Glymour and the Popperians should converge in charging Bayesians with an implicit denial of the value of deep explanatory theories but take as their points of departure opposed positions: Glymour thinks that good explanatory theories by that token justify a correspondingly large claim to belief, and the Popperians assert that such theories merit the lowest possible degree of belief. Whatever their starting points, however, the charge of Glymour, Popper, et al. that Bayesians must in principle undervalue theories is patently false. Perhaps a homely analogy will dispel any lingering doubts that may remain. A jury has always at least two mutually inconsistent hypotheses to consider: that the accused is guilty, and that the accused is not guilty and there is some alternative explanation of the known facts. They wish to determine which, relative to

muddy

those facts, is the more probable hypothesis. Imagine their surprise at being informed that, since they wish to determine the more probable hypothesis relative to the available data, they are thereby committed on their return to the court to announcing that their favoured conclusion is the statement of the factual data they have been given (*see also* Horwich, 1982, p. 132). Scientists, like the court, want information of a specific sort combined with the assurance that it is credible information; and these demands *can*, despite Popper's solemn asseverations to the contrary, simultaneously be met, and the Bayesian theory tells us how.

c THE PRIOR PROBABILITY OF UNIVERSAL HYPOTHESES MUST BE ZERO

Popper, we noted in the previous section, asserts that it is impossible for a hypothesis to possess both high informative content and high probability. In particular, he asserts that the probability of a universal hypothesis must, for logical reasons, be zero (1959a, appendices *vii and *viii). He occasionally remarks (for example, 1959a, p. 381) that the constraints imposed by the probability calculus alone require that the only consistent assignment of a probability to such a hypothesis is zero.

Were Popper correct, then that would be the end of our enterprise in this book, which is to represent the procedures of inductive inference as consistency constraints on the assignments of subjective probabilities. For the truth of Popper's thesis would imply that we could never regard unrestricted universal laws as confirmed by observational or experimental data, since if $P(h) = 0$, then $P(h \mid e) = 0$ also, whatever finite sample data e may consist of. But Popper's thesis is untrue. Even in Carnap's so-called continuum of inductive methods (Carnap, 1952; *see* our discussion in Chapter 3), characterised by the values of a non-negative real parameter λ, one of those methods (corresponding to $\lambda = 0$), assigns, in an obviously consistent way, positive probabilities to a class of strictly universal hypotheses over an infinite domain; and, as we noted in Chapter 3, Hintikka's systems of inductive logic almost invariably assign positive prior probabilities to consistent universal sentences, whether the domain of individuals is finite or infinite.

Popper's arguments for his zero-probability claim are really designed to show something considerably less ambitious than the vastly overstrong thesis that there can be no consistent assignment of a nonzero probability to a universal hypothesis; what they aim at showing, as an examination of his text reveals, is that the so-called *logical* probability of a universal hypotheses must be zero (Popper, 1959, p. 364; a critical examination of these arguments is to be found in Howson, 1973 and 1987). We have already (Chapter 3) discussed the thesis that there is a genuine quantity, the logical probability of a sentence *a*, and concluded that the assumption that there is involves an unacceptably arbitrary, and hence most 'unlogical', degree of apriorism; and a uniform assignment of probability zero to non-tautological universal hypotheses is to our mind no less arbitrary than any other assignment.

Nevertheless, Popper has called our attention to a matter which deserves some comment, namely the fact of the apparent claim to complete generality of much of science and the episodic character of its history, successively punctuated by the demise of great explanatory theories. In view of these facts, we should, it seems, expect all current theory eventually to be overthrown by some new data, new candidates to emerge, become refuted in their turn, and so on ad infinitum. However, it is far from clear that such bleak pessimism really is the lesson taught by the history of science. The mere fact that succeeding extensions of the observational base of science have caused the demise of many an explanatory theory does not demonstrate the appropriateness of total scepticism, nor does it even make it plausible. If up till now I have failed to find the thimble, I do not conclude, and certainly ought not to conclude, that further quest is hopeless. Of course, science is not hunt the thimble, but this does not destroy the point of the analogy, which is that a number of past failures to discover the truth does not by itself imply that one will not one day be successful.

Pessimism on that particular score is certainly not something to which the practitioners of science themselves seem to subscribe. They are not discouraged by the record of others' failures: there is a great deal of biographical and anecdotal evidence which suggests that, on the contrary, some of the most illustrious of scientists not merely invest positive levels of confidence in their theories but, at any rate initially, are frequently prey to the wildest optimism. Watson and Crick quickly were totally convinced that they had discovered the structure of the DNA molecule, to take a well-documented example. This may

not be global physics, but even there the picture is hardly one of unrestrained scepticism. Einstein's confidence in the correctness of his approach notoriously bordered on the hubristic. For example, when, after the reports of the 1919 eclipse expedition, someone asked Einstein what he would have done had the result not been confirmatory of his theory, he replied "Then I would have to pity the dear Lord. The theory is still correct". Even where scepticism supervenes, but accompanied by a determination to persist with the theory—the continuing and fruitful applications of non-relativistic quantum mechanics are a case in point—any subsequent predictive success will usually be attributed to some characteristic feature of that theory even if the particular formulation is thought to be false. But this simply means that it is a suitably modified form of that theory which is regarded as being confirmed. To sum up, there is no evidence that people regard general theories as invariably false, and no evidence that they ought to.

Before we turn to new matters, we must consider the objection that we ourselves are forced, by the way in which we choose to understand Bayesian probabilities, to assign probabilities of zero to universal hypotheses. For we have interpreted these probabilities as implicit assertions about fair betting quotients; to be precise, we regard the assignment of a probability p to h to mean that the individual making that assignment regards p, in the light of his background information, as a fair betting quotient. But if h is universal, then any bet on h might be lost but could not ever be won; consequently, or so it seems, the only fair betting quotient on h is zero, for only the value zero confers no advantage to either side of the bet.

We ourselves in Chapter 3 brought this objection, it may be recalled, against the standard employment of the Dutch Book Argument to show that degrees of belief ought to satisfy the probability calculus. But our use of that same argument is not open to the objection. For we have defined X's degree of belief in the truth of h as what X thinks is the fair betting quotient on h, on the (possibly counterfactual) assumption that the truth-value of h were to be unambiguously decided. That in this sort of case that assumption *is* counterfactual is irrelevant. This is not merely an ad hoc method of evading the problem either, as it might at first sight seem. In fact, it only seems so because of the legacy of Ramsey, Savage, and others, who insisted on behavioural criteria for determining degrees of belief. From the point of view of someone attempting to elicit his or her own strength of belief, on the other hand, introspective thought-experiments, like the one we ourselves invoked, are quite unexceptionable.

■ d PROBABILISTIC INDUCTION IS IMPOSSIBLE

This dramatic claim is made by Popper and David Miller (1983), who also supply what purports to be a rigorous proof of it. Their argument is as follows. According to Bayesian theories of support or confirmation, whether they are subjectively based or not, evidence e supports hypothesis h if and only if $P(h \mid e) > P(h)$. Suppose that h entails e, modulo background information including initial conditions and so forth. Then it is easy to see that e supports h if and only if $P(h) > 0$ and $P(e) < 1$. Suppose that these latter conditions are satisfied also, so that h is (it seems) supported by e. Popper and Miller demonstrate that if in addition $P(h) < P(e)$, then $\sim e \vee h$ is *counter*-supported by e, in the sense that its posterior probability relative to e is *less* than its prior probability (the proof of their result is very straightforward and we shall leave it to the reader to check if they so wish).

This simple theorem of the probability calculus is given a dramatic significance by Popper and Miller. For they claim that $\sim e \vee h$ represents the excess content of h over e, and interpret their result as stating that that excess content is always counter-supported by e. But e may well support h itself; as we saw, it does if h entails e. The Bayesian finds nothing in itself troubling in this breach of what Hempel called the Consequence Condition (that if e confirms h it confirms every consequence of h); he just thinks that the Consequence Condition is false. What does, or ought to trouble him, however, is the explanation of this breach which Popper and Miller provide. For on their interpretation of their formal result, the support e appears to give h is really just the self-support e gives e which is, after all, a part of the content of h. Indeed, if we measure the support $S(h,e)$ of h by e by the simple difference $P(h \mid e) - P(h)$, then it is not difficult to show that $S(h,e) = S(\sim e \vee h,e) + S(e,e)$ when h entails e. All support, conclude Popper and Miller, is really, therefore, self-, or what they call *deductive*, support (since e entails e); the genuinely inductive component $S(\sim e \vee h,e)$ is always negative. So when we think evidence supports a hypothesis, we are, according to Popper and Miller, being misled; it really only supports that part of the hypothesis actually entailed by the evidence, the remainder of the hypothesis being counter-supported.

But Popper's and Miller's argument depends crucially upon identifying the excess content of h relative to e as $\sim e \vee h$, and as Redhead (1985) points out, there is an excellent and simple reason for not doing so. This is that $\sim e \vee h$ is actually a very weak statement (it is entailed by $\sim e$, for example), and h certainly has consequences which are consequences neither of e nor of $\sim e \vee h$: h itself is one of them. This simple fact in our opinion completely demolishes Popper's and Miller's premise that $\sim e \vee h$ "contains everything in h which goes beyond e" (1983, p. 687), and consequently leaves their anti-inductive conclusion quite unsupported.

It is not an adequate rejoinder (nor is it one which Popper and Miller make, incidentally), that since $S(h,e)$ can be split into two additive factors $S(\sim e \vee h,e)$ and $S(e,e)$, this by itself shows that e and $\sim e \vee h$ exhaust the content of h. All that the decomposition shows is that the values of the function $S(h,e)$ can be represented as the sum of the values of the functions $S(\sim e \vee h)$ and $S(e,e)$. The existence of such decomposition certainly does not tell us that the content of h is decomposed into $h \vee \sim e$ and e, since $S(h,e)$ is also decomposable into the sum of the functions $S(h \,\&\, e,e)$ and $S(h \,\&\, \sim e,e)$ (Dunn and Hellman, 1986). Moreover, neither $h \,\&\, e$ nor $h \,\&\, \sim e$ are in general consequences of h, and hence cannot plausibly be regarded as being in the content of h (Dunn and Hellman disagree, incidentally; but see Howson, 1989).

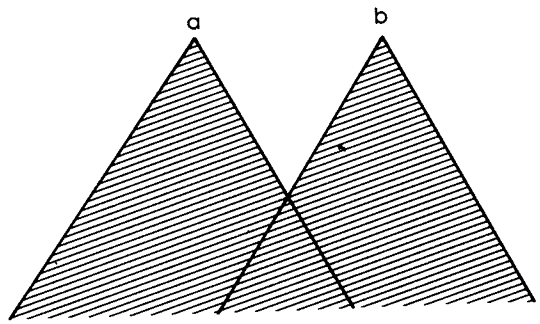
Popper's and Miller's reason for identifying $\sim e \vee h$ as the excess content of h over e is that the conjunction of e and $\sim e \vee h$ is equivalent to h (when h entails e), and that $\sim e \vee h$ and e share no non-tautologous consequences. They argue (1987) that any consequence k of h which is not a consequence either of e or of $\sim e \vee h$ shares non-tautologous consequences with e ($e \vee k$ is one). But this last fact does *not* entail that k is not in the excess content of h over e; it is certainly in the set-theoretic difference of the consequence classes of h and e, and Popper and Miller themselves define the content of a statement to be its set of consequences. The fact that k has consequences in common with e is irrelevant.

But Popper and Miller also employ a numerical measure *ct* of content, *ct(h)* being defined as $P(\sim h)$, where P is what they call a "logical" probability function. And this measure, which is widely endorsed, has the property, as the reader can easily check, that $ct(e) + ct(\sim e \vee h) = ct(h)$, where $h \vdash e$, which would seem to reinforce their claim that e and $\sim e \vee h$ do exhaust the content of h. Hintikka, who discusses the properties of this measure, anticipates Popper and Miller in concluding that $\sim e \vee h$ is correctly identified as the excess content of h over e (1968, p. 313). But we should not be at all impressed by this additivity property of *ct*: all it really means is that *ct* is too coarse-grained to 'notice' that consequences additional to those of e and $\sim e \vee h$ separately are created on conjoining e with $\sim e \vee h$, a fact which is symptomatic of its quite general inability to register the additional content created by the conjunction of two logically independent statements.

Let us briefly justify this last remark. It is easy to show that

$$ct(a) + ct(b) - ct(a \vee b) = ct(a \,\&\, b).$$

Suppose that m is a measure on the set of sets of sentences of the language from which a and b are drawn, and suppose that we identify $ct(c)$ with $mCn(c)$, where $Cn(c)$ is the set of consequences of the sentence c: in other words, suppose that we regard *ct* as measuring consequence classes. $Cn(a \vee b)$ is equal to the intersection of $Cn(a)$ and $Cn(b)$, and it follows that *ct* assigns zero measure to the net increase of consequences created by conjoining a and b. This is best seen graphically in the diagram below, where the baseless cones whose vertices are a and b represent the consequence classes of a and b (the shaded area represents $ct(a \,\&\, b)$).



If one thinks that numerical measures of content are worth investigating at all, then, in view of these observations, one ought to conclude that *ct* is a very *in*adequate and misleading measure of content—though in view of our earlier doubts about whether any useful notion is referred to by the term 'logical probability', any person-independent notion of content based on a probability function seems to be ruled out. Howson and Franklin (1986) investigate numerical content-measures which reflect more faithfully the structure of the underlying consequence classes than does *ct* and show that in this respect more adequate measures are afforded by either $\mathrm{Inf}(h) = -\log P(h)$, the so-called information measure (its expected value is Shannon's entropy), or $\frac{1 - P(h)}{P(h)}$, or the odds against h based on P. Whatever the eventual value of evolving measures of content based on 'suitable' probability functions, we can conclude that the properties of *ct* provide no support for Popper's and Miller's anti-inductivist strictures; nor, as far as we can see, is any to be derived from any other quarter.

e THE PRINCIPAL PRINCIPLE IS INCONSISTENT

David Miller (1966) produces an ingenious argument which appears to demonstrate that the principle which, following Lewis, we have called the Principal Principle, is inconsistent. According to that principle, on which the Bayesian analysis of statistical inference rests, the (subjective) probability that an event described by the sentence a will occur at a particular trial of type T is equal to r, if our data are confined to the information that the physical probability of a, relative to the conditions T, is r. We can write this concisely as the equation

$$\textbf{(1)}\quad P(a_t \mid P^*(a) = r) = r$$

where P is the degree-of-belief probability and P^* the physical probability, and a_t is the statement that a, a generic event-description like 'the coin lands heads 3 times in a sequence of 5 consecutive tosses', is satisfied by the outcome of the particular trial taking place at time t.

Miller's argument is as follows. Let r be $\frac{1}{2}$. Then by (1)

$$P\left(a_t \mid P^*(a) = \frac{1}{2}\right) = \frac{1}{2}.$$

But clearly, $P^*(a) = \frac{1}{2}$ if and only if $P^*(a) = P^*(\sim a)$; and the probability calculus tells us that we can substitute equivalent statements, whence we obtain

(2) $P(a_t \mid P^*(a) = P^*(\sim a)) = \frac{1}{2}$.

However, we can also instantiate (1) thus:

(3) $P(a_t \mid P^*(a) = P^*(\sim a)) = P^*(\sim a)$

and combining (2) and (3) we infer that $P^*(\sim a) = \frac{1}{2}$, which is odd since no factual premise of any kind has been employed in the derivation. While this result may not have the form of an outright contradiction, it very quickly leads to one. For we can repeat the reasoning above with the two substitution instances $P^*(a) = \frac{2}{3}$ *and* $P^*(a) = 2P^*(\sim a)$ instead of $P^*(a) = \frac{1}{2}$ *and* $P^*(a) = P^*(\sim a)$ respectively, whence we infer that $P^*(a) = \frac{2}{3}$; and this is in explicit contradiction to $P^*(a) = \frac{1}{2}$.

Were Miller's derivation formally sound, the consequences for the Bayesian theory of statistical inference would be little short of disastrous; for the Principal Principle provides, as we saw in Chapter 9, the means of evaluating the likelihood terms $P(e \mid h)$ in Bayes's Theorem. Without the principle, Bayes's Theorem would merely contain three undetermined terms, $P(h)$, $P(e)$, and $P(e \mid h)$, where these are either probabilities or probability densities, and would yield no information at all about the value of the posterior probability or probability-density $P(h \mid e)$. The characteristic and often striking properties of the posterior probabilities of statistical hypotheses are due to the behaviour of the likelihood function $g(i,n) = P(e_n \mid h_i)$, where n is the sample size, and $(h_i)_i$ is a family of alternative hypotheses about the value of the physical probability distribution, indexed by a parameter i characterising this distribution, over a sample space one of whose outcomes is e_n.

But, as we also noted in Chapter 9, odds different from those based on the Principal Principle are demonstrably unfair, and this tells us that *something* must be wrong with Miller's clever inconsistency proof. The question is, what? It is certainly not very easy to spot his error, and a considerable number of eminent people have disagreed amongst themselves as to where the error lies. Jeffrey (1970) lists, accompanied by his own, the contemporary analyses of the paradox, though what we believe to be the correct solution was not found until 1979. It is expounded in Howson and Oddie (1979), and we shall reproduce it briefly now.

Miller's error is difficult to spot because it is concealed by the notation which, precisely in consequence of its being well adapted to smooth exposition and development of the theory, does not make explicit all the distinctions which are nevertheless implicit. The erroneous step in Miller's derivation is taking (3) to be a substitution instance of (1). (3) is *not* a substitution instance of (1); it makes a quite different *type* of assertion from (1), and it will help the reader see why if they first turn back to and re-read Chapter 2, section **f.1**, where random-variable-statements are introduced and their meaning discussed. For (1), though it may not look like it, is an equation involving random variables.

This fact is obscured by our tendency to regard $P^*(a)$ as a number. But in the context of a discussion in which '$P^*(a) = r$' is itself a statement assigned a probability value (by the function P) $P^*(a)$ is not a number: it is something which takes a range of possible values—those possible values being, of course, all the real numbers in the closed unit interval. And a quantity which takes different values in different possible states of the world, and over whose values there is a probability distribution, is a random variable. Let us accordingly replace the term $P^*(a)$ in (1) by X, so that (1) becomes

(4) $P(a_t \mid X = r) = r$,

for all r, $0 \leq r \leq 1$. Now recall, from Chapter 2, that we can replace the sentences in (4) by the sets of possible worlds making them true; no formal difference exists between the linguistic and the set-theoretic representation of a class of sentences. So we can rewrite (4) as

(5) $P(M(a_t) \mid M(X = r)) = r$.

Let us now concentrate on $M(X = r)$. Looking back at Chapter 2 again, we see that $M(X = r) = \{w : X(w) = r\}$, where the w's are the members of the outcome space of the stochastic experiment relative to which the distribution P^* is defined. So (5) can be written

(6) $P(M(a_t) \mid \{w : X(w) = r\}) = r$.

But (3) has the form

(7) $P(M(a_t) \mid \{w : X(w) = Y(w)\}) = Y(w)$

where Y is another random variable equal to $1 - X$; Y is of course $P^*(\sim a)$ and $P^*(\sim a) = 1 - P^*(a)$. But it is now obvious that (7) is not a legitimate substitution instance of (5); it contravenes the logical rule that terms involving so-called free variables, like w, must not be substituted into contexts in which those free variable become bound, as the operator $\{w: \ldots\}$ binds $Y(w)$. (*See* Mendelson, 1964, p. 48, for the statement of this rule for logical quantifier operators.) It follows that (3) is not a legitimate substitution instance of (1) and the derivation of Miller's paradoxical conclusion cannot proceed. The Principal Principle is consistent.

■ f HYPOTHESES CANNOT BE SUPPORTED BY EVIDENCE ALREADY KNOWN

This objection is due originally to Clark Glymour (1980), and has since been echoed by many others. Let us quote from Glymour.

> Newton argued for universal gravitation using Kepler's second and third laws, established before the *Principia* was published. The argument that Einstein gave in 1915 for his gravitational field equations was that they explained the anomalous advance of the perihelion of Mercury, established more than half a century earlier.... Old evidence can in fact confirm new theory, but according to Bayesian kinematics it cannot. (p. 86)

By "Bayesian kinematics" Glymour here means simply the prin-
ciple that one's degree of belief in a hypothesis h in the light
of evidence e is equal to one's conditional degree of belief in h
relative to e, which, if one is coherent, is a conditional prob-
ability $P(h \mid e)$. Glymour's thesis that Bayesian kinematics can-
not account for the confirmation of new theories by old facts is
grounded on the undoubted fact that relative to a stock of back-
ground information including e, $P(e)$ is 1, whence $P(e \mid h)$ is 1
also, so that it follows immediately from Bayes's Theorem that
$P(h \mid e) = P(h)$. Thus e does not raise the prior probability of h
and hence, according to the Bayesian, does not confirm it.

Though Glymour's reasoning appears to be sound, it has
the rather strange consequence that *no* data, whether obtained
before or after the hypothesis is proposed, can, within a Baye-
sian theory of confirmation, confirm *any* hypothesis. For even
if the hypothesis h is proposed before evidence e is collected,
then by the time someone comes to do the Bayes's Theorem
calculation, the terms $P(e)$, $P(e \mid h)$ must again be set equal to

1, since by that time e will of course be known and hence in the contemporary stock of background information. But it would be absurd to infer that according to the Bayesian theory e did not support h, and not even the most committed opponent of that theory would claim the theory damaged by this demonstration, for it is clear that the theory has been incorrectly used and that the mistake lies in relativising all the probabilities to the *totality* of current knowledge: they should have been relativised to current knowledge minus e. The reason for the restriction is, of course, that *your current assessment of the support of* h *by* e *measures the extent to which the addition of* e, *to the remainder of what you currently take for granted, would cause a change in your degree of belief in* h.

In other words, once e has become known and you want to assess the support e gives h, the probabilities $P(e \mid h)$, $P(h)$, and $P(e)$ which you will consequently want to evaluate are relativised to the counter-factual knowledge-state in which you still do not know e. Once this is grasped, the solution to the problem of how to deal with the new hypothesis/old evidence problem is obvious in principle, though in some cases difficult to apply in practice: you relativise the probabilities, so far as you are able, to what you currently know minus the data whose current confirmatory capacity vis-à-vis the hypothesis you want to establish.

Glymour considers this reply quite sympathetically, but contends that nevertheless in

> actual historical cases...there is no single counterfactual degree of belief in the evidence ready to hand, for belief in the evidence sentence may have grown gradually—in some cases it may have even waxed, waned, and waxed again. (p. 88)

He cites as an example the data on the perihelion of Mercury; there were different values obtained for this, over a period of several decades, by different methods, and employing mathematical techniques sometimes without rigorous justification. Glymour contrasts this situation with the results of tossing a coin a specified number of times, where he thinks it does make sense to talk of the probability of that outcome, as if it had not yet occurred. But in the case of Mercury's estimated perihelion advance, "there is no single event, like the coin-flipping, that makes the perihelion anomaly virtually certain" (Glymour, 1980, p. 88).

But whether there is as much epistemic warrant for the data in 1915 about the magnitude of Mercury's perihelion advance as there is about the number of heads we have just observed in a sample of a hundred tosses of a coin is beside the point. About some data we may be more tentative, about other data less. The Bayesian theory we are proposing is a theory of inference from data; we say nothing about whether it is correct to accept the data, or even whether your commitment to the data is absolute. It may not be, and you may be foolish to repose in it the confidence you actually do. The Bayesian theory of support is a theory of how the acceptance as true of some evidential statement affects your belief in some hypothesis. How you came to accept the truth of the evidence, and whether you are correct in accepting it as true, are matters which, from the point of view of the theory, are simply irrelevant. Glymour's disquisition on the frailty of much scientific data is therefore, however valuable in its own right, beside the point of evaluating the adequacy of the Bayesian theory of inference.

The same is true of his subsequent discussion of the lack of any *general* means of computing these degrees of belief. He considers and rejects several candidates, and concludes that it may not be some "old result that confirms a new theory, but rather the new discovery that the new theory entails (and thus explains) the old [result]" (Glymour, 1980, p. 92). This observation prompts Glymour to propose a new, quasi-Bayesian criterion for old evidence *e* to be taken as confirming a new theory *h*, namely that

(8) $P(h \mid e \,\&\, (h \vdash e)) > P(h \mid e)$

where the probability calculus is weakened appropriately, by replacing the conditions on axioms 2 and 3 by 'if it is known that *t* is a tautology then . . .', and 'if it is known that $a \vdash \sim b$ then . . .' (we have modified Glymour's notation in (8) slightly and omitted explicit reference to background information). This emendation of the classical theory is sympathetically endorsed by Niiniluoto (1983), and is further examined by Garber (1983), though the same idea seems first to have been proposed and developed by I. J. Good (starting with Good, 1948), who calls the resulting notion of probability "dynamic", or "evolving" probability.

Let us consider these issues in turn: first, the alleged absence of any general rule for computing *P(e)* in the counterfactual way suggested. We can ignore this discussion because the sort of Bayesianism we are advocating is not regarded as a source of rules for computing all the probabilities in Bayes's Theorem. In particular, we are under no obligation to legislate concerning the methods people adopt for assigning prior probabilities. These are supposed merely to characterise their beliefs subject to the sole constraint of consistency with the probability calculus. In the context of this particular discussion, all we are concerned with is that people are capable in many cases of determining, possibly only very roughly, to what extent they think a piece of data likely relative to a stock of residual background information.

What of Glymour's suggestion that it is not simply the knowledge of *e*, but in addition the discovery that *h* explains *e*, on which one bases the conclusion that *h* is confirmed by *e*? This is undoubtedly true; the question, which Glymour thinks is to be answered in the affirmative, is whether this should prompt any change in the formalism of the Bayesian theory if the latter is to reflect it adequately. On the standard account, the knowledge that *h* entails *e* is discharged in computing $P(e \mid h)$ to be 1, and allowance for the recognition of entailment relations is made generally in this way. It is true, however, that this has often been deemed unrealistic, because it implies, to take a dramatic example, that the degree of belief of any coherent individual in the four-colour conjecture, conditional on the truth of certain other mathematical statements, is 1, and always was 1 from the time the conjecture was first formulated; whereas, of course, it was exceedingly difficult, and took a great deal of time (including computer time) to prove that result.

But we are not asserting that everybody ought to be consistent, nor that people actually are invariably consistent. The foundation of subjective Bayesianism is probabilistic consistency, admittedly, but this is an ideal which is not always attainable, and when complex logical entailments are involved, we should not expect that ideal to be realised in more than a small number of cases, if at all. It is certainly no part of our use of the theory as an explanatory tool that people are invariably consistent any more than it is the case that somebody who wished to provide an account of deductive inference is committed to the view that people either ought to be or are invariably deductively consistent. All we claim is that when people recognise or are apprised of deductive relationships between hypothesis and evidence, they often draw conclusions about levels of support in accordance with those determined within the Bayesian theory by the same initial probabilities.

This means that Glymour's non-classical condition (8) does not after all make for a markedly, if at all, more realistic theory of Bayesian inference. Nor is it necessary to introduce that confirmation condition (8), in the context of a weakened probability calculus, in order to give an adequate Bayesian account of how old evidence can confirm new theories. On this account, recall, the probabilities which are used to determine support are relativised not to one's total background information, but to what that state of background information would be were one not yet to know the data in question. While Glymour's objections to this proposal are, we have argued, unconvincing, there remain two further objections which we must now consider.

The first is that relativisation of all the probabilities to what is strictly a fictitious state of background information is simply ad hoc: it is a device which avoids the otherwise embarrassing necessity of setting $P(e)$ and $P(e \mid h)$ equal to 1, but it does so at the cost of being in conflict with core Bayesian principles. But a little reflection should convince the reader that this charge is untrue. Core Bayesian principles simply state the conditions—obedience to the probability calculus—for a set of degrees of belief, relative to a stock of background information, to determine a corresponding set of odds which are not demonstrably unfair. There is absolutely nothing in this which asserts that in computing levels of support, one's subjective probabilities must define degrees of belief relative to the totality of one's current knowledge. On the contrary; as we pointed out earlier, the support of h by e is gauged according to the effect which a knowledge of e *would* now have on one's degree of belief in h, on the (counter-factual) supposition that one does not yet know e.

The second objection occurs in a paper by Campbell and Vinci (1983). Their argument is in essentials as follows. Suppose first of all that h predicts, relative to suitable initial conditions, an event e, and the experiment designed to elicit e if h is true has not yet been performed. Suppose also that relative to current background information, $P(e)$ is high. The experiment is duly performed and e is observed. The support of h by e is inconsiderable, because of course $P(e)$ is high. Now suppose that $P(e)$ is high precisely because e describes the same sort of effect as has already been observed to occur in contexts which are thought to be strongly analogous to those in which e's occurrence is predicted by h. Let the conjunction of all this past data be e'. Suppose that h was proposed well after e' was known, and that h also 'predicts' e' in the relevant circumstances, although it was not deliberately designed to do so (the reason for this rider will only become fully apparent when we come to section *g*). Finally, suppose that relative to background information minus e, $P(e')$ is low. According to our analysis, the support of h by e' is considerable, unlike its support by e.

Campbell and Vinci flesh out the picture by taking h to be statistical thermodynamics and e and e' to describe examples of Brownian motion in different liquids. *Their* conclusion, for which incidentally they produce no argument but seem to take as intuitively justified, is that the support of h by e ought to be no different from its support by e', which they agree is high (p. 323). But this conclusion is far from intuitively clear, and indeed seems to us to be, on the contrary, intuitively quite wrong: what we are seeing here is, after all, the phenomenon of diminishing support by repeated occurrences of the same type of effect. Initially (e') the support is high; later (e) it is not. This 'intuition' (for want of a better word) is reinforced by our analysis. There is not much more that we can say. We have given reasons, which seem to us good ones, for a conclusion disputed by Campbell and Vinci on grounds which presumably are supposed to be intuitive, but to us seem anything but.

g HYPOTHESES ARE NOT SUPPORTED BY DATA THEY WERE CONSTRUCTED TO EXPLAIN

There is one feature of the Campbell–Vinci example which may well cause the reader disquiet; indeed, it may appear that the discussion of that example implicitly refutes the Bayesian theory of support. The feature we are referring to is the innocent-looking clause stipulating that h was not designed to accommodate e'. The attentive reader may have noted that according to our analysis of the situation, h would be supported by e' to exactly the same extent *whether this condition were satisfied or not*. Nothing in the method of computing this can take account of whether h were deliberately constructed with the explanation of e' in mind or not.

We have already discussed in Chapter 4 the claim that a hypothesis designed to fit some piece of data is not supported by it to as great an extent as one which also fits the data but accidentally, so to speak. We argued there that this claim was false; and the Bayesian theory of support is certainly inconsistent with it. But there are arguments for the view, and these both sound convincing and also number among their subscribers many if not most contemporary philosophers of science. We shall examine these arguments now and show that their plausibility vanishes on closer inspection.

Suppose that data e are employed as a constraint in constructing a deterministic hypothesis h in such a way that h is made to entail e if suitable initial conditions are met. It seems fairly clear that h is not under risk of falsification by e; and in consequence of this fact it is argued that e cannot support h. The argument is a popular one; thus Giere writes

> If the known facts were used in constructing the model and were thus built into the resulting hypothesis . . . then the fit between these facts and the hypothesis provides no evidence that the hypothesis is true [since] these facts had no chance of refuting the hypothesis. (1984, p. 161)

Glymour (1980, p. 41) voices a substantially identical opinion, as does Zahar, in a slightly more elaborate way (1983, p. 245). But the argument is quite fallacious. First, note that it is simply false that any *fact* has a "chance" of refuting anything. If e is a factual statement and h a hypothesis, then e either refutes h or it does not, and it does or does not whether h was designed to explain or embody e or not. The confusion here is a confusion of a random variable (in effect, the experimental set-up E) with one of its possible outcomes, e. What has the chance of refuting h, if at least one of its possible outcomes is inconsistent with h, is E, and only E. And when it is E rather than e which is recognised to have this chance, the argument collapses, for even if h was constructed from one particular outcome of E, it is in general logically possible that E could have produced an outcome e' which is inconsistent with h.

Suppose that h entails e modulo initial conditions which obtain when the experiment E is performed. It is interesting to note that Giere's corrected principle—that if E stands no

chance of refuting h, then h cannot be supported by the outcome e of E either—follows directly from Bayes's Theorem itself if we are allowed to render "chance" as "probability", for to say that E has no chance of refuting h entails that $\sim e$ has no chance of occurring, which translates into the condition that $P(e) = 1$. Since $P(e \mid h) = 1$, it follows that $P(h \mid e) = P(e)$ and that the support for h from e is zero. But that $P(e) = 1$ certainly does not follow from the fact that h was constructed to explain e.

This conclusion, however, is implicitly contested by both Giere (1984) and Redhead (1986), for they argue that if h is deliberately constructed to explain e, then the probability of e given that h is false is equal to 1. But from $P(e \mid \sim h) = 1$ it easily follows, given that $P(e \mid h) = 1$, that $P(e) = 1$, since $P(e) = P(e \mid h)P(h) + P(e \mid \sim h)P(\sim h) = P(h) + P(\sim h) = 1$. How do these authors manage to conclude what seems, then, clearly false, that $P(e \mid \sim h) = 1$ when h is constructed to explain e? Let us take Giere first. He considers Mendel's simple one-factor, two-allele model of inheritance in his pea plants, which Giere contends was constructed to explain the two-to-one ratio of tall to dwarf plants in the second filial generation after crossing true-breeding tall and dwarf in the parental generation, and then mating the tall offspring. Giere asserts that "fitting this case was . . . a necessary requirement for any model to be seriously entertained. So there seems no way [in which the data could be improbable given the negation of Mendel's hypothesis]" (1983, p. 118).

Giere provides no argument for this claim, however: on the contrary, he appears to take it as self-evident. But it is far from self-evident that Mendel's data would not be improbable were his own explanation of them to be false; indeed, as that was the *only* explanation which seemed plausible to Mendel, its falsity would presumably render those data, were they assumed to be still conjectural, relatively improbable as far as he was concerned. And this, as we have seen, is sufficient for a Bayesian to be able to explain the undoubted fact that Mendel himself took his data to be strongly confirmatory of his model. Giere has not justified his thesis—nor indeed could he—that $P(e \mid \sim h) = 1$ when h has been designed to explain e.

Let us now look at Redhead's argument for that claim. The curious thing about the argument he offers is that it purports to be a Bayesian argument; but the Bayesian position seems quite opposed to any such conclusion, as we have seen. But let us see what Redhead says. He commences by casting Giere's premise, that in seeking to explain e one is making the explanation of e a necessary condition for any model to be entertained, into the more explicitly Bayesian condition that the desire to explain e acts as a "filter" upon the set of all hypotheses, allowing a non-zero prior probability only to those which, relative to a suitable set of auxiliary hypotheses and initial conditions, call them a, entail e. It certainly follows from this "filter" condition that $P(e \mid h) = 1$, since

$$P(e \mid \sim h) = \frac{\Sigma P(e \mid h_i)P(h_i)}{P(\sim h)},$$

where the sum is over all the mutually exclusive hypotheses h_i whose disjunction is equivalent, with probability 1, to $\sim h$ (the disjunction may include what Shimony (1970) calls a "catch all" hypothesis, simply equivalent to the negation of the disjunction of all the remaining h_i and h). Redhead's filter condition ensures that all the h_i in the sum are such that $P(e \mid h_i) = 1$, and hence that $P(e \mid \sim h) = \frac{\Sigma P(h_i)}{\Sigma P(h_i)} = 1$. But this, of course, implies Giere's conclusion, that h receives null support from e, since $P(e \mid \sim h) = P(e \mid h) = 1$.

One obvious fault of Redhead's assumption that, in setting out to construct an explanation of e one is, in effect, assigning a positive prior probability only to those hypotheses which entail e relative to a, is that it is inconsistent. For a tautology does not explain e, yet its probability must be one. Even if the filtering is restricted to some partition $\{h_i\}$ (and it is not clear that anyone determines such a partition is attempting to explain e), the condition that only those members have positive probability which entail e is still far too strong. For it yields exactly the same, null-support, result for hypotheses proposed independently of the data e as it does for those deliberately designed to explain e. Suppose h is designed to explain e, modulo a. It follows from the filter condition that $P(e)$ equals 1, since $P(e) = P(e \mid h)P(h) + P(e \mid \sim h)P(\sim h) = P(h) + P(\sim h) = 1$. But since $P(e) = 1$ it also must be the case that $P(e \mid \sim h') = 1$, where h' is any hypothesis which entails e modulo a. It follows that $P(h' \mid e) = P(h')$, so that h' is not supported by e either. The unwelcome strength of Redhead's filter condition stems from its implicitly making $P(e) = 1$. As we have already pointed out, however, $P(e)$ is not equal to unity just

because e is a known effect we wish explained.

We conclude that attempts to show that data which hypotheses have been deliberately designed to entail, as opposed to independently predicting, do not support those hypotheses fail. On the contrary, the condition for support, that $\frac{P(e \mid \sim h)}{P(e \mid h)}$ be small, may be perfectly well satisfied in many such cases. We have already (Chapter 4) looked at some simple examples, though no harm will come from repeating them here. (i) a box is known to contain a number k of red and blue balls. We want to know the proportion of red balls. We take the balls out one by one and note their colour, until the box empty. We note that there are r red balls. We formulate the hypothesis h that the proportion of red balls originally in the urn is $\frac{r}{k}$. h is constructed from the data, and nobody would presumably deny that it is supported by the data since, together with some uncontentious background information, the data entail h and vice versa; thus $\frac{P(e \mid \sim h)}{P(e \mid h)}$ is zero.

If this example is thought too trivial and 'unscientific', consider (ii): we have the same box, but use a different experiment to discover the proportion of red balls. We take from the box a sequence of n balls, noting the colour of each, replacing it and shaking the box. Suppose that s of the balls we observe in this way are red. We formulate the hypothesis h' that the proportion of red balls in the box is $\frac{s}{n} \pm \epsilon$, where ϵ is a suitable real number. h' too would be regarded by most people as a hypothesis well supported by the data. Here, however, $P(e \mid h')$ is not only not equal to 1: it will be very small indeed if n is greater than 20, say, because it will be obtained by applying the background model of a binomial distribution with probability parameter p in the interval $\frac{s}{n} \pm \epsilon$. But while $P(e \mid h')$ is small, $P(e \mid \sim h')$ is in general very much smaller. (If we take the binomial model as *the* model of the experiment, with a uniform prior probability distribution over the values of the binomial parameter, then $P(e \mid \sim h')$ can straightforwardly be computed using the theorem of total probabilities; even if the prior distribution is not uniform, then the same method will give a good approximation for sufficiently large n.) The reference to balls in a box can be dispensed with in evaluating the significance of the example in (ii), incidentally, which is modelled in a great variety of estimably scientific experiments.

It might be objected that in regarding the hypotheses in (i) and (ii) as well supported by the data from which they were calculated, we are relying on background information containing well-defined models of the experiment: the data, in fact, merely perform the function of specifying parameters in those underlying models which are themselves taken to be independently very well-supported, to the point of their truth being taken for granted. This is true, but not damaging. In evaluating support we always and necessarily employ *some* background information which we take pretty much or even completely for granted. And this background information will always be analogous to a model with undetermined parameters, in that it will leave open a (more-or-less indefinite) range of alternative hypotheses about the structure of some experimental process. Of course, the data obtained from that experimental source will by no means uniquely determine which hypothesis is correct, given that background information; but then the data in (ii) above failed to fix the parameter uniquely either. Of course, also, cases of parameter fixing are very special examples of data-determining hypotheses: the space of possibilities is clearly defined, for one thing, with usually a simple mathematical structure (it is often an interval of real numbers). But these features do not at all affect the validity of our general conclusion, which is that in appropriate circumstances some data might both act as a constraint on the construction of hypotheses and simultaneously support those hypotheses; and the appropriateness of the circumstances can be characterised, albeit in general terms, as a function of the magnitude of the likelihood ratio $\frac{P(e \mid \sim h)}{P(e \mid h)}$.

■ h PREDICTION OR ACCOMMODATION?

There is a fairly ancient, many-sided debate about confirmation, one side of which asserts that hypotheses constructed deliberately to accommodate data e are never supported to the same (positive) extent by e as hypotheses which independently predict e. An extreme version of this view is that such data never support the corresponding hypotheses. The burden of the

previous few paragraphs is that this extreme thesis is untenable. We have also argued (Chapter 4, section **j**) that the less extreme view (that had h independently predicted e, then it would have obtained more support from e than had it been constructed with e as an explicit constraint) is also false in general. However, we shall now exhibit some, not altogether atypical, circumstances in which the accommodating hypothesis receives at most as much support as the independently predicting one.

Suppose, for example, that h possesses an undetermined parameter a. An experiment is conducted to determine the value of a relative to the assumption that h is true. Its outcome is e. Let the resulting hypothesis, namely h together with the computed (or estimated) value of a, be h', and suppose that h', conjoined with the initial conditions of the experiment, entails e (it does not have to, in general). Now suppose that another hypothesis h'' had been formulated and proposed before that experiment had been performed, but together with the same initial conditions, it too entails e. Here we have two hypotheses, in other words, one (h') which has been made to accommodate e, and the other (h'') which independently predicts e. (It may, of course, be the case that fixing the parameter in h at some particular value enables h', the adjusted hypothesis, to explain e, but at the cost of making other predictions which are known to be false. The de Sitter modification of Poincaré's Lorentz-invariant gravitational theory is a case in point, where the parameter adjusted to yield the correct value of Mercury's annual precession caused the theory to be incorrect about the bending of light, among other things.)

Finally, suppose that prior to that experiment the prior probability of h'' is at least as great as that of h. Since we are now talking about relative magnitudes of support, we want some way of measuring supports on some numerical scale. We have already taken the function $S(\cdot,e) = P(\cdot \mid e) - P(\cdot)$, where the relevant hypothesis replaces the dot, as a natural measure which locates degree of support within the interval $[-1,1]$, and we shall use S in what follows. Where the initial conditions of the experiment are regarded as being part of the general background information, we have that $S(h',e) = \frac{P(h')[1 - P(e)]}{P(e)}$ and similarly that $S(h'',e) = \frac{P(h'')[1 - P(e)]}{P(e)}$. So the ratio of the supports of h' and h'' is just the ratio of their prior probabilities. However, h' logically entails h, so that $P(h') \leq P(h)$, and by assumption $P(h) \leq P(h'')$. Hence $S(h',e) \leq S(h'',e)$.

We said that the condition for that inequality, namely that $P(h) \leq P(h'')$, is not too atypical. It often happens, for example, that one scientific theory predicts an effect e which serves only to fix a parameter-value in a rival theory. We are tempted to say that the first theory gets more support from e than does the second, even if the latter also 'predicts' e in those circumstances once the parameter has been fixed. But it is important to be clear about *why* we make this judgment. It is not because independent prediction *always* confers more support than accommodation, as an influential tradition commencing with Leibniz and including Whewell has claimed. Support depends on prior probability, as the support-function S makes clear, and a completely incredible theory will not in general be regarded as being supported whatever it predicts, or how it did so (in the limit we have a contradiction which predicts everything and is supported by nothing, for example). We make the judgment that the independently predicting theory h'' usually gets more support than the adjusted hypothesis h' precisely because h and h'' are rivals, and hence can be presumed to have comparable prior probabilities. And if those probabilities are equal, within the limits of imprecision which usually attends such judgments, then the inequality above will be valid.

These conclusions depend on the fact that the prior probability of h, the hypothesis with free parameters, determines the maximum support which is gained by h' from the data which fixed the values of the parameters: as we saw, $S(h',e) \leq \dfrac{P(h)[1 - P(e)]}{P(e)}$. This is a result of considerable significance. For it tells us, among other things, that a consequence of regarding the introduction of those parameters as *merely* an ad hoc way of accommodating the data, then the support of the resulting determinate hypothesis will certainly be inconsiderable. The plausibility of the thesis that predictions always glean more support than accommodations rests, we suspect, on nothing more than invalidly generalising from this special case in which the thesis is true (this point is made forcefully by Nickles, 1985, p. 200).

We can illustrate these remarks with an example we used earlier, in Chapter 4, section **k,** that of the new 'law' of free fall

$$(**)\quad s = g(t) + f(t)(t - t_1)(t - t_1) \ldots (t - t_n),$$

where $g(t)$ is Galileo's law and $t_1, \ldots, t_n$ are the time instants at which observations were made of the corresponding values of s, and $f(t)$ is some arbitrary function of t which, we shall assume, is nowhere zero.

Clearly, (**) is obtained from the hypothesis that there are nonzero values of $a_1, \ldots, a_n$ such that

$$(***)\quad s = g(t) + f(t)(t - a_1)(t - a_2) \ldots (t - a_n)$$

and the n observations; the latter uniquely determine the values $a_1 = t_1, \ldots, a_n = t_n$ of the parameters a_i once the functional form (***) is given. Intuitively (let us suppose that we are in the late sixteenth century), the prior probability of the parametric model (***) is zero or at most negligible in comparison with that of the Galilean law $g(t)$, because the parameters a_i and the function $f(t)$ are introduced purely ad hoc, for no reason whatever except that it is known that they will fit the observations.

(**) and (***) and the Galilean law itself are therefore instances of the hypotheses h', h, and h'' in our discussion above. We infer that the support of (**) by the n observations is smaller than that of the Galilean law, simply as a result of the distribution of prior probabilities over the parametric model (***) and Galileo's law. (**) is an extreme case of curve-fitting by adjustment of parameters whose introduction lacks any theoretical justification whatever.

To sum up: it is not true that data which are used as explicit constraints on the construction of explanatory hypotheses are thereby precluded from being counted in support of the resulting hypotheses. Nor is it true that a hypothesis which independently predicts the data gets more support from it than one which is constructed in order to predict it. The widespread belief to the contrary is an illicit extrapolation from those cases where the data-generated hypothesis is based on a theoretical model whose sole raison d'être is that it accommodates the observations, usually by being endowed with an appropriately large number of adjustable parameters. In these cases, where the theoretical model has no independent justification, support is withheld. The Bayesian theory here as elsewhere only articulates the feelings of the practitioners themselves: the economist Edgeworth, for example, expressed his reserve about Karl Pearson's fitting his (Pearson's) family of probability-density curves to the data, in the pointed question (Edgeworth 1895, p. 511) "what weight should be attached to this correspondence by one who does not perceive any theoretical reason for those formulas". Kepler fitted ellipses to Tycho's data for planetary orbits, but only after he had found independent reasons for that type of orbit. Nearer to home we find Kitcher castigating some sociobiologists' parameter adjustment on the ground that "the model gives absolutely no insight into the reasons behind the periodicity [the adjusted parameter] . . . the choice of a periodic function for the probability bears no relation to any psychological mechanisms" (1985, p. 375). And so on.

I THE PRINCIPLE OF CONDITIONALISATION, AND BAYESIAN LEARNING

Somebody who has degrees of belief $P(h)$, $P(e)$, and $P(h \,\&\, e)$ in the truth of the sentences h, e, and $h \,\&\, e$ thereby has, on pain of inconsistency, as we saw in Chapter 3, a degree of belief $P(h \mid e)$ in h, conditional on e's being true, where $P(h \mid e) = \dfrac{P(h \,\&\, e)}{P(e)}$. If e does turn out to be true, then the degree of belief of this person, again on pain of inconsistency, in h unconditionally becomes $P'(h) = P(h \mid e)$. This is the Principle of Conditionalisation whose validity we also proved in Chapter 3. It has, however, become a focus of critical attention in the past few years, and its status disputed. Hacking, Kyburg, Levi, and many others claim that the principle requires a justification independent of that of the axioms of the probability calculus, as we noted in Chapter 3; but we also observed there that the claim is false. However, there are some apparent objections to the principle, and these we must examine now.

One objection is that the Principle of Conditionalisation is the only mechanism in the Bayesian theory for learning by experience. Since the prior probability distributions which enter into the Bayes's Theorem expression of $P(h \mid e)$ themselves reflect what van Fraassen (1980) calls "the deliverances of experience", it would seem that they can achieve a Bayesian explanation only if they themselves are posterior probabilities, relative to some anterior reception of data and some yet prior probabilities; and so on, until an ultimately prior distribution is reached, prior to all empirical experience, far back in the history of the organism, at the dawn of its cognitive life. To explain how current empirical data affect current belief then seems to entail not only the reconstruction of the agent's successive acts of conditionalisation—a daunting if not practically impossible task—but also the characterisation of a state of primal ignorance; and we have seen in Chapter 3 that attempting to characterise primal ignorance in terms of some absolutely neutral prior distribution is a pretty hopeless task.

But none of this need worry us. There is nothing in the account we have given which commits us to the thesis that all change of belief takes place via Bayesian conditionalisation. In fact, that account implicitly contradicts such a thesis, for apart from anything else the data e appearing in the conditional probabilities are given exogenously: e is, for want of a better word, simply 'known'. The reader, aware of the fallibility of practically all 'deliverances of experience', may regard this admission as entailing a no less damagingly unrealistic account of inductive inference. We shall argue shortly that this is not so. But first things first. We are proposing a theory of inference; in particular, a theory in which from two inputs, e and a belief distribution, an output belief-distribution is generated by the Principle of Conditionalisation. Since we are not claiming that *all* belief distributions are obtained by conditionalisation, we are not committed to explaining the provenance of the input belief-distribution.

A more serious objection is that the data input e is simply taken as given. Many people have regarded this as an embarrassment for the Bayesian theory, because it has seemed to them that 'given' here is synonymous with 'certain'. Keynes distinguished between indirect, merely probable knowledge, arising from conditionalising on direct, or certain knowledge (1921, pp. 10–20), and in doing so set the terms of the subsequent debate to the extent that a quarter of a century later C. I. Lewis could remark that "If anything is probable, then something must be certain" (Lewis 1946, p. 186). Lewis's observation is, or so it would appear, fully endorsed by the Principle of Conditionalisation, since by that principle, $P'(e) = P(e \mid e) = 1$. However, as we conceded earlier, our knowledge of the world is not simply decomposable into two kinds, that which is infallible and that which is conjectural. If no one else, then Descartes should have made us aware that practically nothing is certain.

One well-known way of accommodating this objection, which yields the Principle of Conditionalisation as a special case, is due to Richard Jeffrey. Jeffrey's model for belief change as a result of experiential inputs ("probability kinematics") does not involve the ascription of any probability at all, let alone probability 1, to them. In the simplest possible case, we revise, as a result of some experience, our personal probability of one hypothesis h from $P(h)$ to $P'(h)$. How should this affect the probabilities of the various other hypotheses we contemplate? According to Jeffrey (1983, p. 169), if a is any other hypothesis then

(9) $P'(a) = P(a \mid h)P'(h) + P(a \mid {\sim}h)P'({\sim}h)$

Clearly, if $P'(h) = 1$, then we obtain the ordinary rule of conditionalisation, $P'(a) = P(a \mid h)$, so that "conditionalisation is a limiting case of the present more general method of assimilating uncertain evidence, and the case of conditionalisation is approximated more and more closely as the probability [of h] approaches 1" (p. 171). The case where more than one hypothesis has its probability exogenously altered is a straightforward generalisation of (9) (ibid.).

Both the Principle of Conditionalisation and Jeffrey's rule emerge as special cases in the theory of belief change presented in Williams (1980), in which the posterior distribution P of belief over a class of mutually exclusive hypotheses h_i is determined as that distribution which minimizes what Williams and others call the information in P relative to a prior distribution P^0, subject to whatever constraints are imposed as a consequence of some experiential input. Introduced by Hobson (1971) as the unique quantity satisfying some intuitively plausible desiderata for measures of relative information, the information in P relative to P^0, $I(P,P^0)$, is defined to be equal to

$$\sum_i P(h_i)\log\left(\frac{P(h_i)}{P^0(h_i)}\right).$$

I is intended to measure something like the probability-relevant magnitude of the information whose acquisition changes P^0 to P. Thus I is zero when P and P^0 are the same distribution (it is always nonnegative), and it becomes large without bound if P places an event close to 1 which P^0 places close to 0. It is also not difficult to see that the function P minimizing I, subject to the condition that for some statement a in the domain of P, $P(a) = 1$, is such that $P(h_i) = P^0(h_i \mid a)$ for every i. If instead we take the constraint, as in the Jeffrey situation, merely to

be that $P(a)$ is some number between 1 and 0, then minimizing information subject to that constraint yields

$$P(h_i) = P^0(h_i \mid a)P(a) + P^0(h_i \mid {\sim}a)P({\sim}a),$$

that is to say, we obtain Jeffrey's rule (Williams, 1980 p. 136).

A great deal has been written about the status of Jeffrey's rule. It should be emphasised that it is not the only way of generalising the Principle of Conditionalisation to contexts in which evidence is more or less uncertain: there is an infinity of ways. Nor is it justified by the sorts of consistency constraint that we have invoked to justify the probability axioms and the rule of conditionalisation (Armendt, 1980, provides however a Dutch Book argument for it, but only given certain other conditions). Nor, as Levi points out, is it clear how "to distinguish those initial shifts [from $P(h)$ to $P'(h)$] for which [the agent] has no justification from those [e.g. from $P(a)$ to $P'(a)$] which he can justify via an appeal to the initial shifts" (1967, p. 204). As far as Williams's rule is concerned, similar considerations apply. In particular, Williams offers no real argument why the appropriate posterior probability is the one which minimises I relative to the prior P^0. In summary, both Jeffrey's rule and Williams's generalisation are undoubtedly interesting and valuable developments in the Bayesian account, but as yet they remain speculative developments; their status as an extension of core principles needs more in the way of justification than they have yet got. For this reason nothing in our account will depend on their acceptance.

However, there remains the problem posed by the general fallibility of evidence statements and the fact that in the ordinary Bayesian account they are assigned posterior probability 1. There is a prima facie conflict here, as we noted above, which only some Jeffrey-type relaxation of the probability 1 condition can, it seems, avert. Despite appearances, however, there is really no conflict. In our account there is nothing that demands that what is taken as data in one inductive inference cannot be regarded as problematic in a later one. Assigning probability 1 to some data on one occasion does not mean that on all subsequent occasions it need be assigned probability 1. Levi (1967) has argued similarly, and persuasively, for just this point in his discussion of Jeffrey's rule. He also points out that it is a mistake to regard the ascription of probability 1 to a hypothesis as equivalent to the assertion that that statement is infallible. All that the ascription of probability one to *e* entails, in our and Levi's view, is that the agent takes *e* to be true in the light of his current experience. It does not follow that at some future occasion he might not have equally compelling reasons to regard *e* as false: *e* remains corrigible, in other words, but may quite reasonably, given appropriate background data, be currently assigned probability 1. Levi sums up the position very nicely:

> propositions accorded probability one are liable to be false. . . . The ramifications of this approach do admittedly stand in need of further examination. But the position is frankly fallibilistic. Empirical propositions can justifiably be believed and, indeed, admitted into evidence even though it is possible that they are false. (1967, p. 209)

For these reasons, then, we feel that the fact of the fallibility of data poses no threat to the use of the rule of conditionalisation. It is time to move on.

■ j THE PROBLEM OF SUBJECTIVISM

Possibly the most serious of all objections made against the subjective Bayesian theory is that it is simply too subjective. Fisher, in his remark which we quoted in Chapter 3, section **c**, that results concerning the measurement of belief "are useless for scientific purposes", summed up what many thought and still think to be a crucial objection. Science is objective to the extent that the procedures of inference in science are. But if those procedures reflect purely personal beliefs to a greater or lesser extent, as they appear to do if they are constrained only to follow Bayes's Theorem, with no condition other than mere consistency being imposed on the forms of the priors, then the inductive conclusions so generated will also reflect those purely personal opinions. Echoing Fisher, E. T. Jaynes claims that

> the most elementary requirement of consistency demands that two persons with the same relevant prior information should assign the same prior probabilities. Personalistic doctrine makes no attempt to meet this requirement . . . the notion of personalistic probability belongs to the field of psychology and has no place in applied statistics. Or, to state this more constructively, objectivity requires that a statistical analysis should make use, not of anybody's personal opinions, but rather the specific factual data on which those opinions are based. (1968, quoted in Rosenkrantz 1977, p. 53)

Alas, neither Jaynes nor his followers are able to live up to his ideal; nor is it possible in principle that they could. No prior distribution reflects only factual data unmixed with anybody's opinions. This is true simply because no prior distribution reflects only factual data, so any given prior distribution will reflect an opinion of some sort. Thus Rosenkrantz, an enthusiastic supporter of Jaynes, defends the uniform prior distributions that tend to arise within Jaynes's theory by pointing out an analogy with current cosmological practice:

> Steady-state cosmologists, to take one of myriad instances, start off by assuming the laws of physics are the same in temporally and spatially remote regions of the universe. This, they urge, is surely the simplest assumption. But it is more than that. To assume that different laws obtained a billion years ago would be entirely arbitrary; it would be to import knowledge we do not in fact possess. (Rosenkrantz, 1977, p. 54)

But we don't know that the laws were the same either. Rosenkrantz has failed to see that *any* assumption 'imports knowledge', the assumption that things were essentially the same just as much as the assumption that they were not. So it is with a uniform, or any other prior distribution, as we argued at length (possibly ad nauseam) in Chapter 3. Because it is important, however, let us repeat the fundamental fact once more. *No prior probability or probability-density distribution expresses merely the available factual data; it inevitably expresses some sort of opinion about the possibilities consistent with the data.* Even a uniform prior distribution is defined only relatively to some partition of these possibilities: we can always find another with respect to which the distribution is as biased as you like—or don't like.

Jaynes's objective priors do not exist. But it does not follow, as he, and Fisher, and people too numerous to mention (among them Bayesian personalists) think it follows, that without objective priors the Bayesian theory is constrained to be a record merely of the whims of individual psychology. That quite fallacious inference has been possibly more damaging to rational methodology than any other. We have pressed the analogy with deductive logic, and we shall press it again. Deductive logic is the theory (though it might be more accurate to say 'theories') of deductively valid inferences from premisses whose truth-values are exogenously given. Inductive logic—which is how we regard the subjective Bayesian theory—is the theory of inference from some exogenously given data and prior distribution of belief to a posterior distribution. Both logics assign categorical status to certain distinguished types of statement (tautologies, for example, are necessarily true and necessarily have probability 1). Most importantly, as far as the canons of inference are concerned, neither logic allows freedom to individual discretion: both are quite *im*personal and objective.

Moreover the subjective Bayesian theory *does,* as we have seen, incorporate Jaynes's requirement that "two persons with the same relevant prior information" assign the same prior probabilities, but it does so asymptotically, as their data garnered from experience grow without bound. Even then, as we point out in Chapter 11, it characteristically does not take all that much sample data to diminish the different distributions to the point where they are practically identical. Experience is allowed to dominate prior beliefs, in other words, though in a controlled way; disagreement is not eradicated at once, which seems entirely natural, but its effect usually falls off quickly. What more could anybody—reasonably—want?

This consequence of the Bayesian theory, namely the tendency of experience to reduce disagreement, is usually brought out as the sole line of defence against the charge of idiosyncratic subjectivism. While it is important, indeed very important, it is not the sole and should not even be the principal defence, however. For the charge, as we have attempted to show, is quite misconceived. It arises from a widespread failure to see the subjective Bayesian theory for what it is, a theory of inference. And as such, it is unimpeachably objective: though its subject matter, degrees of belief, is subjective, the rules of consistency imposed on them are not at all.

■ k SIMPLICITY

What, though, it may be asked, about invoking a criterion of simplicity as a method of constraining prior distributions—a method which has the virtue both of being objective and of conforming to actual scientific practice? How many times, after all, have we read scientists' claims that it was the great sim-

plicity of such and such a theory which made them repose such high initial confidence in it and remain convinced of its truth long after adverse empirical evidence would have seen off a less intrinsically appealing theory? Why not, therefore, incorporate a criterion of simplicity explicitly as a constraint on ranking prior probabilities?

But it is not at all clear actually how this advice should be followed. Simplicity turns out to be a highly elusive concept, which has so far resisted all attempts to characterise it in any uncontroversial way. Some people maintain that simplicity resides in an organic unity exemplified by the fundamental principles of the simple theory. Others say that it resides in the fewness of the adjustable parameters which the theory introduces. Yet others say it resides in the ease with which computations can be done within the theory. But all these notions, where any clear sense may be made of them, appear to be independent of each other. And it is notoriously difficult, moreover, to make any clear sense of them. Even such an apparently perspicuous notion as paucity of independent parameters is, on inspection, far from ambiguous. Newton's theory, for example, might be thought to possess very few undetermined parameters—some people claim that it contains only one, the gravitational constant. But Newton's theory applied in, say, the kinetic theory of gases, contains of the order, even in the simplest applications, of 10^{23} undetermined parameters, and when further degrees of freedom are added to these ideal models, the number rises proportionately.

The reader must forgive us if we do not add to the vast and inconclusive literature on simplicity. We think that its inconclusive nature is symptomatic of ambiguity and obscurity in the very notion itself. But our main reason for concluding the discussion here is implicit in what we have already said about the nature of our enterprise. We wish to lay down no indefensible a priori principles as universal standards, any more than we should expect a deductive logician to add to the list of logical axioms a statement of the Principle of the Uniformity of Nature, say, simply because it seems to have been adopted implictly or explicitly by so many people. This is not to say that we do not see the Bayesian theory as highly explanatory: we do, but we would take people's apparent sensitivity to considerations of simplicity, in whatever particular form such considerations might take, merely as a true factual statement about one of the components *for them* of a high prior probability.

But here we are obliged to correct a well-known claim of Popper's. According to him, if one takes the paucity-of-parameters analysis of simplicity, then it "*contradicts* the laws of the calculus of probability" to assign greater prior probabilities to the simpler of any two hypotheses (1959, p. 381, his italics). This claim occurs in his discussion of Jeffreys and Wrinch's (1921) so-called simplicity ordering, according to which a pair of rival hypotheses would be assigned probabilities in the way Popper contends to be impossible (*see also* Jeffreys 1961, pp. 45–48 for his further development of the idea of a simplicity ordering). But Popper's claim is incorrect. There is no inconsistency at all in assigning a higher probability to a hypothesis which asserts only that a trajectory is a curve of degree 2 than to that which asserts only that it is of degree 3. Popper's own arguments for his false claim are in fact based on much more than the probability calculus; they are based either on the Classical Theory of probability, or on the principle that the 'more easily testable' a hypothesis is—and one which is simpler, in this particular sense, is more easily testable—the lower should be its prior probability. We have argued at length in Chapter 3 that the use of the Classical Theory of probability to generate prior probability distributions is quite arbitrary. Popper's 'more easily testable = less a priori probable' equation is equally arbitrary: there can be no grounds for assuming that, because fewer independent observations are required to test h_1 than h_2, h_1 is less likely to be true than h_2. There is no more reason to believe this than there is to believe (as Jeffreys did) that h_1 is more likely to be true than h_2 (for a fuller discussion of these issues see Hesse 1974, p. 226–227, and Howson, 1973, 1987, and 1988).

■ I PEOPLE ARE NOT BAYESIANS

In their summary of an influential piece of empirical work, Kahneman and Tversky deliver themselves of the following judgment:

> The view has been expressed . . . that man, by and large, follows the correct Bayesian rule, but fails to appreciate the full impact of evidence [they cite W. Edwards 1968], and is therefore conservative.
>
> The usefulness of the normative Bayesian approach to the analysis and the modeling of subjective probability depends primarily not on the accuracy of the subjective estimates, but rather on whether the model captures the essential determinants of the judgment process. The research discussed in this paper suggests that it does not. . . . In his evaluation of evidence, man is apparently not a conservative Bayesian: he is not Bayesian at all. (Kahneman and Tversky [1972] p. 46).

It has been the burden of the foregoing chapters that a Bayesian theory is capable of explaining standard modes of scientific inference where other theories are not. Yet the empirical studies Kahneman and Tversky refer to are taken by these authors to indicate very strongly that people do not use Bayesian reasoning where the Bayesian theory appears to say that they should. Kahneman and Tversky consider the objection to their conclusion, that obtaining posterior probabilities even relative to simple Bernoulli trials requires a degree of practice and sophistication which most people do not possess; they dismiss it, however, citing as evidence that the sorts of departures from the model they find seem to be always of the same type, and independent of the mathematical sophistication of the subjects, and even of their acquaintance or not with basic probability theory. How can we—indeed, can we—continue to regard the Bayesian theory as explanatory in the face of adverse evidence like this?

However, apportioning the blame between a central hypothesis and the various auxiliaries required to test it, when adverse results are obtained, is known to be a problematic affair (this is the Duhem problem, discussed in Chapter 4, section **h**). There is no justification for the definitive conclusion arrived at by Kahneman and Tversky that people are not in agreement with Bayesian precepts in the way they process data. All the cases which are supposed to show this depend on the assumption of the subjects' tacit acceptance of some type or other of probability model which the testers think appropriate. If the subjects had chosen some different model, for whatever reason, then the conclusions to be drawn may be entirely different. Indeed, the author of one of the most celebrated studies to which Kahneman and Tversky refer questions their conclusion for that reason: the subjects' perception of the appropriate model may not have been the testers' (Phillips, p. 1983, p. 531).

However, we should be surprised if on every occasion subjects were apparently to employ impeccable Bayesian reasoning, even in the circumstances that they themselves were to regard Bayesian procedures as canonical. It is, after all, human to err, and sometimes to err in very distinctive and persistent ways. It is instructive to compare the situation described by Kahneman and Tversky with a rather striking and very uniform result (one of the present authors has tested it himself on a group of American freshman and sophomore students) of an experiment, devised by P. C. Wason (1966) to test subjects' performance of, on the face of it, a simple deductive task. Four cards are placed flat on a table. Each card has an integer between 1 and 4 inclusive printed on one face and a letter on the other. The uppermost faces of the cards are

E	K	4	7

and the subjects are asked to name those cards, and only those cards, which need to be turned over in order to determine whether the statement, 'if a card has a vowel on one side, then it has an even number on the other', is true. Wason discovered that the vast majority of his subjects indicated either the pair of cards E and 4, or only the card 4. The correct answer is, of course, the pair E and 7.

This empirical result has proved to be remarkably persistent:

> Time after time our subjects fall into error. Even some professional logicians have been known to err in an embarrassing fashion, and only the rare individual takes us by surprise and gets it right. It is impossible to predict who he will be. This is all very puzzling... (Wason and Johnson-Laird, 1972, p. 173)

Puzzled Wason and Johnson-Laird may be, but about one thing they are certainly clear: these subjects did get the answer wrong. Moreover, even the subjects themselves eventually agreed on that. Now this observation has an obvious relevance to Kahneman's and Tversky's dramatic claim, made in the light of evidence anaologous to Wason's, that we are not Bayesians. Wason has shown, by this and other empirical studies, that we are not consistently deductive logicians in practice. But he has not shown, nor did he claim to have shown, that we are not deductive logicians in some other important sense. For we ourselves nevertheless constructed those deductive standards and consciously attempt to meet them, even though we sometimes fail, and in some cases nearly always fail. By the same token, it is not prejudicial to the conjecture that *what we ourselves*

take to be correct inductive reasoning is Bayesian in character that there should be observable and sometimes systematic deviations from Bayesian precepts.

■ m CONCLUSION

One of the reasons why one expects deductive reasoning to exercise a more-or-less widely felt and obeyed constraint on the way people reason is because it is truth preserving. Probabilistic reasoning also possesses a characteristic which authorises it to exercise no less a regulatory function: its rules, as we observed in Chapter 3, are broken on pain of committing inconsistency. It is, we suggest, for this reason that divergence from the norm set by the probability calculus is also regarded as deviant. Certainly ever since people chose to express their uncertainty in terms of the odds they thought fair, they have felt themselves explicitly constrained by the axioms of the probability calculus, and while it was not until this century that it was explicitly proved that obedience to the calculus is a necessary condition for fairness, there can be little doubt that that result was taken for granted.

The discovery of the probability calculus, together with the usual formula connecting (fair) odds and probabilities, can now be seen to be part of the great scientific renaissance of the seventeenth century. The probability calculus became the foundation of a mathematical theory of uncertainty, of enormous potential scope and power, which simultaneously generated a quantitative logic of inductive inference and bound together the new mathematical concept of probability with another developed at about the same time, utility, to produce a theory of rational action. The mathematicians of the eighteenth century, and to a lesser extent the nineteenth, divided their time between developing the new physics and extending the probability calculus and the theory of inductive inference and rational decision based on it: among these pioneers, Huyghens, James and Daniel Bernoulli, Laplace, and Poisson stand out as pre-eminent.

On the way, however, paradoxes began to appear in the programme, mostly connected with the Principle of Indifference but also—as a criterion of rational action—with the principle of expected utility. These problems, especially those within the theory of probability itself, seemed at one time, in the early years of this century, so intractable that many people, like Fisher and Popper as we have seen, wrote off the account of probability on which the programme was based. But they were wrong: in the middle years of this century, shortly after Fisher and Popper penned their obituaries, secure foundations were finally laid. Von Neumann and Morgenstern put utility theory on a consistent basis, and Ramsey and de Finetti realised that an adequate theory of epistemic probability can dispense with pseudo-objective principles like that of Indifference without giving up its claim to impose quite objective standards of consistency in reasoning involving such probabilities. The probabilities might be personal, but the constraints imposed on them by the condition of consistency are certainly not—a distinction still not widely grasped even today, and whose failure to be appreciated continues to vitiate so much contemporary discussion.

We have written this book in an attempt to convince believers in 'objective' standards in science that there is nothing subjective in the Bayesian theory *as a theory of inference*: its canons of inductive reasoning are quite impartial and objective. We want this simple truth to be more widely appreciated, and not only this one. Equally, we want to demonstrate to those same people that this is the *only* theory which is adequate to the task of placing inductive inference on a sound foundation. The rival claims of the other approaches we have examined in the previous chapters are quite spurious and often do not withstand even a cursory inspection. We hope that we have been at least partially successful in achieving these objectives: the final judgment must, however, as always, be the reader's.

BIBLIOGRAPHY

Altman, D. G., Gore, S. M., Gardner, M. J., and Popcock, S. J. 1983. 'Statistical Guidelines for Contributors to Medical Journals', *British Medical Journal*, vol. 286, 1489–493.

Armendt, B. 1980. 'Is there a Dutch Book Argument for Probability Kinematics?', *Philosophy of Science*, vol. 47, 583–89.

Bacon, F. 1620. *Novum Organum*. In *The Works of Francis Bacon*, vol. 4, J. Spedding, R. L. Ellis, and D. D. Heath, eds. 1857–58, London: Longman and Company.

Babbage, C. 1827. 'Notice respecting some Errors common to many Tables of Logarithms', *Memoirs of the Astronomical Society*, vol. 3, 65–67.

Barnett, V. 1973. *Comparative Statistical Inference*. Chichester: Wiley.

Bayes, T. 1763. 'An essay towards solving a problem in the doctrine of chances', *Philosophical Transactions of the Royal Society*, vol. 53, 370–418. Reprinted with a biographical note by G. A. Barnard in *Biometrika*, 1958, vol. 45, 293–315.

Bernoulli, J. 1713. *Ars Conjectandi*. Basiliae.

Bolzano, B. 1837. *Wissenschaftstheorie* (English translation by Rolf George published under the title *Theory of Science*, Blackwell, 1972).

Bowden, B. V. 1953. 'A Brief History of Computation', in *Faster than Thought*, edited by B. V. Bowden. London: Pitman Publishing.

Campbell, R. and Vinci, T. 1983. 'Novel Confirmation', *British Journal for the Philosophy of Science*, vol. 34, 315–341.

Carnap, R. 1947. 'On the Applications of Inductive Logic', *Philosophy and Phenomenological Research*, vol. 8, 133–148.

———. 1950. *Logical Foundations of Probability*. Chicago: University of Chicago Press.

———. 1952. *The Continuum of Inductive Methods*. Chicago: University of Chicago Press.

Carnap, R. and Jeffrey, R., eds. 1971. *Studies in Inductive Logic and Probability*. Berkeley: University of California Press.

Church, A. 1940. 'On the Concept of a Random Sequence', *Bulletin of the American Mathematical Society*, vol. 46, 130–35.

Cochran, W. G. 1952. 'The χ^2 Test of Goodness of Fit', *Annals of Mathematical Statistics*, vol. 23, 315–345.

———. 1954. 'Some Methods for Strengthening the Common χ^2 Tests', *Biometrics*, vol. 10, 417–451.

Cournot, A. A. 1843. *Exposition de la Théorie des Chances et des Probabilités*. Paris.