CHAPTER 7

Bayesian Versus Non-Bayesian Approaches

In this chapter we shall consider how, by attributing positive probabilities to hypotheses in the manner described in Chapter 2, one can account for many of the characteristic features of scientific practice, particularly as they relate to deterministic theories.

a THE BAYESIAN NOTION OF CONFIRMATION

Information gathered in the course of observation is often considered to have a bearing on the acceptability of a theory or hypothesis (we use the terms interchangeably), either by confirming it or by disconfirming it. Such information may either derive from casual observation or, more commonly, from experiments deliberately contrived in the hope of obtaining relevant evidence. The idea that evidence may count for or against a theory, or be neutral towards it, is a central feature of scientific inference, and the Bayesian account will clearly need to start with a suitable interpretation of these concepts.

Fortunately, there is a suitable and very natural interpretation, for if $P(h)$ measures your belief in a hypothesis when you do not know the evidence e, and $P(h \mid e)$ is the corresponding measure when you do, e surely confirms h when the latter exceeds the former. So we shall adopt the following as our definitions:

e **confirms or supports** h when $P(h \mid e) > P(h)$

e **disconfirms or undermines** h when $P(h \mid e) < P(h)$

e **is neutral with respect to** h when $P(h \mid e) = P(h)$

One might reasonably take $P(h \mid e) - P(h)$ as measuring the degree of e's support for h, though other measures have been suggested (e.g., Good, 1950); disagreements on this score will not need to be settled in this book. We shall refer, in the usual way, to $P(h)$ as 'the prior probability of h' and to $P(h \mid e)$ as h's 'posterior probability' relative to, or in the light of, e. The reasons for this terminology are obvious, but it ought to be noted that the terms have a meaning only in relation to evidence: as Lindley (1970, p. 38) put it, "[t]oday's posterior distribution is tomorrow's prior". It should be remembered too that all the probabilities are evaluated in relation to accepted background knowledge.

b THE APPLICATION OF BAYES'S THEOREM

Bayes's Theorem relates the posterior probability of a hypothesis, $P(h \mid e)$, to the terms $P(h)$, $P(e \mid h)$, and $P(e)$. Hence, knowing the values of these last three terms, it is possible to determine whether e confirms h, and, more importantly, to calculate $P(h \mid e)$. In practice, of course, the various probabilities may only be known rather imprecisely; we shall have more to say about this practical aspect of the question later.

The dependence of the posterior probability on the three terms referred to above is reflected in three striking phenomena of scientific inference. First, other things being equal, the extent to which evidence e confirms a hypothesis h increases with the likelihood of h on e, that is to say, with $P(e \mid h)$. At one extreme, where e refutes h, $P(e \mid h) = 0$; hence, disconfirmation is at a maximum. The greatest confirmation is produced, for a given $P(e)$, when $P(e \mid h) = 1$, which will be met in practice when h logically entails e. Statistical hypotheses, which will be dealt with in Parts III, IV, and V of this book, are more substantially confirmed the higher the value of $P(e \mid h)$.

Secondly, the posterior probability of a hypothesis depends on its prior probability, a dependence sometimes discernible in scientific attitudes to ad hoc hypotheses and in frequently expressed preferences for the simpler of two hypotheses. As we shall see, scientists always discriminate, in advance of any experimentation, between theories they regard as more-or-less credible (and, so, worthy of attention) and others.

Thirdly, the power of e to confirm h depends on $P(e)$, that is to say, on the probability of e when it is not assumed that h is

true (which, of course, is not the same as assuming h to be false). This dependence is reflected in the scientific intuition that the more surprising the evidence, the greater its confirming power. However, $P(e) = P(e \mid h)P(h) + P(e \mid \sim h)P(\sim h)$ (as we showed in Chapter 2, section **c.3**), so that really, the posterior probability of h depends on the three basic quantities $P(h)$, $P(e \mid h)$, and $P(e \mid \sim h)$.

We shall deal in greater detail with each of these facets of inductive reasoning in the course of this chapter.

c FALSIFYING HYPOTHESES

A characteristic pattern of scientific inference is the refutation of a theory, when one of a theory's empirical consequences has been shown to be false in an experiment. As we saw, this kind of reasoning, with its straightforward and unimpeachable logical structure, exercised such an influence on Popper that he made it the centrepiece of his scientific philosophy.

Although the Bayesian approach was not conceived specifically with this aspect of scientific reasoning in view, it has a ready explanation for it. The explanation relies on the fact that if, relative to background knowledge, a hypothesis h entails a consequence e, then (relative to the same background knowledge) $P(h \mid \sim e) = 0$. Interpreted in the Bayesian fashion, this means that h is maximally disconfirmed when it is refuted. Moreover, as we should expect, once a theory is refuted, no further evidence can ever confirm it, unless the refuting evidence or some portion of the background assumptions is revoked. (The straightforward proofs of these claims are suggested as an exercise.)

d CHECKING A CONSEQUENCE

A standard method of investigating a deterministic hypothesis is to draw out some of its logical consequences, relative to a stock of background knowledge, and check whether they are true or not. For instance, the General Theory of Relativity was confirmed by establishing that light is deflected when it passes near the sun, as the theory predicts. It is easy to show, by means of Bayes's Theorem, why and under what circumstances a theory is confirmed by its consequences.

If h entails e, then, as may be simply shown, $P(e \mid h) = 1$. Hence, from Bayes's Theorem: $P(e \mid h) = \frac{P(h)}{P(e)}$. Thus, if $0 < P(e) < 1$, and if $P(h) > 0$, then $P(h \mid e) > P(h)$. It follows that any evidence whose probability is neither of the extreme values must confirm every hypothesis with a non-zero probability of which it is a logical consequence.

Succeeding confirmations must eventually diminish in force, for the theory has an upper limit of probability beyond which no amount of evidence can push it. This too follows from Bayes's Theorem. Suppose $e_1, e_2, \ldots, e_n, \ldots$ are consequences of h. Then Bayes's Theorem asserts that

$$P(h \mid e_1 \,\&\, e_2 \,\&\, \ldots \,\&\, e_n) = \frac{P(h)}{P(h \mid e_1 \,\&\, e_2 \,\&\, \ldots \,\&\, e_n)}.$$

Now

$$P(e_1 \,\&\, e_2 \,\&\, \ldots \,\&\, e_n) = P(e_1)P(e_2 \,\&\, \ldots \,\&\, e_n \mid e_1)$$

and

$$P(e_2 \,\&\, \ldots \,\&\, e_n \mid e_1) = P(e_2 \mid e_1)P(e_3 \,\&\, \ldots \,\&\, e_n \mid e_1 \,\&\, e_2).$$

Thus, in general,

$$P(e_1 \,\&\, e_2 \,\&\, \ldots \,\&\, e_n) = P(e_1)P(e_2 \mid e_1) \ldots P(e_n \mid e_1 \,\&\, \ldots \,\&\, e_{n-1}).$$

Hence,

$$P(h \mid e_1 \,\&\, e_2 \,\&\, \ldots \,\&\, e_n) = \frac{P(h)}{P(e_1)P(e_2 \mid e_1) \ldots P(e_n \mid e_1 \,\&\, \ldots \,\&\, e_{n-1})}.$$

Provided $P(h) > 0$, the term $P(e_n \mid e_1 \,\&\, \ldots \,\&\, e_{n-1})$ must tend to 1 as n increases. If it did not, the posterior probability of h would at some point exceed 1, which is impossible (Jeffreys, 1961, pp. 43–44). This explains why it is not sensible to test a hypothesis indefinitely, though without more detailed information on the individual's belief-structure, in particular regarding the values of $P(e_n \mid e_1 \,\&\, \ldots \,\&\, e_{n-1})$, one could not know the precise point beyond which further predictions of the hypothesis were sufficiently probable not to be worth examining.

Specific categories of a theory's consequences also have a restricted capacity to confirm (Urbach, 1981). Suppose h is the theory under discussion and that h_r is a substantial restriction of that theory. A substantial restriction of Newton's theory might, for example, express the idea that freely falling bodies near the earth descend with a constant acceleration or that the period and length of a pendulum are related by the familiar formula. Since h entails h_r, $P(h) \leq P(h_r)$—(*see* Chapter 2, section **c.3**)—and if h_r is much less speculative than its progenitor, it will often be significantly more probable.

Now consider a series of predictions derived from h, but which also follow from h_r. If the predictions are verified, they may confirm both theories, whose posterior probabilities are given by Bayes's Theorem, thus:

$$P(h \mid e_1 \,\&\, e_2 \,\&\, \ldots \,\&\, e_n) = \frac{P(h)}{P(e_1 \,\&\, e_2 \,\&\, \ldots \,\&\, e_n)}$$

and

$$P(h_r \mid e_1 \,\&\, e_2 \,\&\, \ldots \,\&\, e_n) = \frac{P(h_r)}{P(e_1 \,\&\, e_2 \,\&\, \ldots \,\&\, e_n)}.$$

Combining these two equations to eliminate the common denominator, one obtains

$$P(h \mid e_1 \,\&\, e_2 \,\&\, \ldots \,\&\, e_n) = \frac{P(h)}{P(h_r)} P(h_r \mid e_1 \,\&\, e_2 \,\&\, \ldots \,\&\, e_n).$$

Since the maximum value of the last probability term in this equation is 1, it follows that however many predictions of h_r have been verified, the main theory, h, can never acquire a posterior probability in excess of $\frac{P(h)}{P(h_r)}$. Hence, the type of evidence characterised by entailment from h_r may well be limited in its capacity to confirm h.

This result explains the familiar phenomenon that repetitions of a particular experiment often confirm a general theory only to a limited extent, for the predictions verified by means of a given kind of experiment (that is, an experiment designed to a specified pattern) do normally follow from and confirm a much-restricted version of the predicting theory. When an experiment's capacity to generate confirming evidence has been exhausted through repetition, further support for h would have to be sought from other experiments, experiments whose outcomes were predicted by different parts of h.

The arguments and explanations in this section rely on the possibility that evidence already accumulated from an experiment may increase the probability of further performances of the experiment producing similar results. Such a possibility is denied by Popperians on the grounds that the probabilities involved are subjective. How then do they explain the fact, attested by every scientist, that by repeating some experiment, one eventually (usually quickly) exhausts its capacity to confirm a given hypothesis? Alan Musgrave (1975) attempted an explanation designed on Popperian lines. He claimed that after a certain, unspecified number of repetitions of an experiment, the scientist would form a generalisation to the effect that whenever the experiment was performed, it would yield a similar result. Musgrave then proposed that the generalisation should be entered into 'background knowledge'. Relative to this newly augmented background knowledge, the experiment is certain to produce a similar result at its next performance. Musgrave then appealed to the principle that evidence confirms a hypothesis in proportion to the difference between its probability relative to the hypothesis together with background knowledge and its probability relative to background knowledge alone. That is, the degree to which e confirms h is proportional to $P(e \mid h \,\&\, b) - P(e \mid b)$, b being background knowledge. Musgrave then inferred that even if the experiment did produce the expected result when next performed, the hypothesis would receive no new confirmation. Watkins (1984, p. 297) has endorsed this account.

A number of decisive objections may be raised against it, though. First, as we shall show in the next section, although it seems to be a fact and is an essential constituent of Bayesian reasoning, there is no basis in Popperian methodology for confirmation to depend on the probability of the evidence; Popper simply invoked the principle ad hoc. Secondly, Musgrave's suggestion takes no account of the fact that particular experimental results may be generalised in infinitely many ways. This is a substantial objection, since different generalisations give rise to different expectations about the outcomes of future experiments. Musgrave's account is incomplete without a rule to specify in each case the appropriate generalisation that should be formulated and adopted, and it is hard to imagine how such a rule could be justified within the confines of Popperian philosophy. Finally, the decision to designate the generalisation background knowledge, with the consequent effect on our evaluation of other theories and on our future conduct regarding, for example, whether to repeat certain experiments, is comprehensible only if we have invested some confidence in the theory. But then Musgrave's account tacitly calls on the same kind of inductive considerations as it was designed to circumvent, so its aim is defeated.

■ e THE PROBABILITY OF THE EVIDENCE

The degree to which h is confirmed by e depends, according to Bayesian theory, on the extent to which $P(e \mid h)$ exceeds $P(e)$. An equivalent way of putting this is to say that confirmation is correlated with the difference between $P(e \mid h)$ and $P(e \mid \sim h)$, that is, with how much more probable the evidence is if the hypothesis is true than if it is false. This is obvious from the third form of Bayes's Theorem (*see* Chapter 2):

$$\frac{P(h \mid e)}{P(h)} = \frac{1}{P(h) + \frac{P(e \mid \sim h)}{P(e \mid h)} P(\sim h)}.$$

These facts are reflected in the everyday experience that information that is particularly unexpected or surprising, unless some hypothesis is assumed to be true, supports that hypothesis with particular force. Thus, if a soothsayer predicts that you will meet a dark stranger sometime and you do, your faith in his powers of precognition would not be much enhanced: you would probably continue to think his predictions were just the result of guesswork. However, if the prediction also gave the correct number of hairs on the head of that stranger, your previous scepticism would no doubt be severely shaken.

Cox (1961, p. 92) illustrated this point with an incident in *Macbeth*. The three witches, using their special brand of divination, predicted to Macbeth that he would soon become both Thane of Cawdor and King of Scotland. Macbeth finds both these prognostications almost impossible to believe:

> By Sinel's death, I know I am Thane of Glamis,
> But how of Cawdor?
> The Thane of Cawdor lives, a prosperous gentleman,
> And to be King stands not within the prospect of belief,
> No more than to be Cawdor.

But a short time later he learns that the Thane of Cawdor prospered no longer, was in fact dead, and that he, Macbeth, has succeeded to the title. As a result, Macbeth's attitude to the witches' powers is entirely altered, and he comes to believe in their other predictions and in their ability to foresee the future.

The following, more scientific, example was used by Jevons (1874, vol. 1, pp. 278–79) to illustrate the dependence of confirmation on the improbability of the evidence. The distinguished scientist Charles Babbage examined numerous logarithmic tables published over two centuries in various parts of the world. He was interested in whether they derived from the same source or had been worked out independently. Babbage (1827) found the same six errors in all but two and drew the "irresistible" conclusion that, apart from these two, all the tables originated in a common source.

Babbage's reasoning was interpreted by Jevons roughly as follows. The theory t_1, which says of some pair of logarithmic tables that they shared a common origin, is moderately likely in view of the immense amount of labour needed to compile such tables ab initio, and for a number of other reasons. The alternative, independence theory might take a variety of forms, each attributing different probabilities to the occurrence of errors in various positions in the table. The only one of these which seems at all likely would assign each place an equal probability of exhibiting an error and would, moreover, regard those errors as more-or-less independent. Call this theory t_2 and let e^i be the evidence of i common errors in the tables. The posterior probability of t_1 is inversely proportional to $P(e^i)$, which, under the assumption of only two rival hypotheses, can be expressed as $P(e^i) = P(e^i \mid t_1)\,P(t_1) + P(e^i \mid t_2)P(t_2)$. (This is the theorem of total probability—*see* Chapter 2, section **c.3**.) Since t_1 entails e^i, $P(e^i) = P(t_1) + P(e^i \mid t_2)P(t_2)$. The

quantity $P(e^i \mid t_2)$ clearly decreases with increasing i. Hence $P(e^i)$ diminishes and approaches $P(t_1)$, as i increases; and so e^i becomes increasingly powerful evidence for t_1, a result which agrees with scientific intuition.

In fact, scientists seem to regard a few shared mistakes in different mathematical tables as so strongly indicative of a common source that at least one compiler of such tables attempted to protect his copyright by deliberately incorporating three minor errors "as a trap for would-be plagiarists" (L. J. Comrie, quoted by Bowden, 1953, p. 4).

The relationship between how surprising a piece of evidence is on background assumptions and its power to confirm a hypothesis is a natural consequence of Bayesian theory and was not deliberately built in. On the other hand, methodologies that eschew probabilistic assessments of hypotheses seem constitutionally incapable of accounting for the phenomenon. Such approaches would need to be able, first, to discriminate between items of evidence on grounds other than their deductive or probabilistic relation to a hypothesis. And having established such a basis for discriminating, they must show a connection with confirmation. The objectivist school has more-or-less dodged this challenge. An exception is Popper. In tackling the problem, he moved partway towards Bayesianism; however, the concessions he made were insufficient. Thus Popper conceded that, in regard to confirmation, the significant quantities are $P(e \mid h)$ and $P(e)$, and as we have already reported, he even measured the amount of confirmation (or "corroboration", to use Popper's preferred term) which e confers on h by the difference between these quantities (Popper, 1959a, appendix *ix).

But Popper never stated explicitly what he meant by the probability of evidence. On the one hand, he would never have allowed it to have a subjective connotation, for that would have compromised the supposed objectivity of science; on the other hand, he never worked out what objective significance the term could have. His writings suggest that he had in mind some purely logical notion of probability, but as we saw in Chapter 4, there is no adequate account of logical probability. Popper also never explained satisfactorily why a hypothesis benefits from improbable evidence or, to put the objection another way, he failed to provide a foundation in non-Bayesian terms for the

Bayesian confirmation function which he appropriated. (For a discussion and decisive criticism of Popper's account, see Grünbaum, 1976.)

The Bayesian position has recently been misunderstood to imply that if some evidence is known, then it cannot support any hypothesis, on the grounds that known evidence must have unit probability. That the objection is based on a misunderstanding is shown in Chapter 15, where a number of other criticisms of the Bayesian approach will be rebutted.

f THE RAVENS PARADOX

That evidence supports a hypothesis more the greater the ratio $\frac{P(e \mid h)}{P(e)}$ scotches a famous puzzle first posed by Hempel (1945) and known as the *Paradox of Confirmation* or sometimes as the *Ravens Paradox*. It was called a paradox because its premisses were regarded as extremely plausible, despite their counter-intuitive, or in some versions contradictory, implications, and the reference to ravens stems from the paradigm hypothesis ('All ravens are black') which is frequently used to expound the problem. The difficulty arises from three assumptions about confirmation. They are as follows:

1. Hypotheses of the form 'All *R*s are *B*' are confirmed by the evidence of something that is both *R* and *B*. For example, 'All ravens are black' is confirmed by the observation of a black raven. (Hempel called this Nicod's condition, after the philosopher Jean Nicod.)
2. Logically equivalent hypotheses are confirmed by the same evidence. (This is the Equivalence condition.)
3. Evidence of some object not being *R* does not confirm 'All *R*s are *B*'.

We shall describe an object that is both black and a raven with the term *RB*. Similarly, a non-black, non-raven will be denoted $\bar{R}\,\bar{B}$. A contradiction arises for the following reasons: an *RB* confirms 'All *R*s are *B*', on account of the Nicod condition. According to the Equivalence condition, it also confirms 'All non-*B*s are non-*R*s', since the two hypotheses are

logically equivalent. But contradicting this, the third condition implies that *RB* does not confirm 'All non-*B*s are non-*R*s'.

The contradiction may be avoided by revoking the third condition, as is sometimes done. (We shall note later another reason for not holding on to it.) However, although the remaining conditions are compatible, they have a consequence which many philosophers have regarded as blatantly false, namely, that by observing a non-black, non-raven (say, a red herring or a white shoe) one confirms the hypothesis that all ravens are black. (The argument is this: 'All non-*B*s are non-*R*' is equivalent to 'All *R*s are *B*'; according to the Nicod condition, the first is confirmed by $\bar{R}\,\bar{B}$; hence, by the Equivalence condition, so is the second.)

If non-black, non-ravens support the raven hypothesis, this seems to imply the paradoxical result that one could investigate that and other generalisations of a similar form just as well by observing white paper and red ink from the comfort of one's writing desk as by studying ravens on the wing. However, this would be a non sequitur. For the fact that *RB* and $\bar{R}\,\bar{B}$ both confirm a hypothesis does not imply that they do so with equal force. Once it is recognised that confirmation is a matter of degree, the conclusion is no longer so counter-intuitive, because it is compatible with $\bar{R}\,\bar{B}$ confirming 'All *R*s are *B*', but to a minuscule and negligible degree.

Indeed, most people do have a strong intuition that an *RB* confirms the ravens hypothesis (h) more than an $\bar{R}\,\bar{B}$. We can appreciate why that might be by consulting Bayes's Theorem as it applies to the two types of datum:

$$\frac{P(h \mid RB)}{P(h)} = \frac{P(RB \mid h)}{P(RB)} \quad \& \quad \frac{P(h \mid \bar{R}\,\bar{B})}{P(h)} = \frac{P(\bar{R}\,\bar{B} \mid h)}{P(\bar{R}\,\bar{B})}$$

These expressions can be simplified. First, $P(RB \mid h) = P(B \mid h \,\&\, R)P(R \mid h) = P(R \mid h) = P(R)$. We arrived at the last equality by assuming that whether some arbitrary object is a raven is independent of the truth of h, which seems plausible to us, at any rate as a good approximation, though Horwich (1982, p. 59) thinks it has no plausibility. By similar reasoning, $P(\bar{R}\,\bar{B} \mid h) = P(\bar{B} \mid h) = P(\bar{B})$. Also $P(RB) = P(B \mid R)P(R)$, and $P(B \mid R) = \sum P(B \mid R \,\&\, \theta)P(\theta \mid R) =$ (assuming independence between θ and R) $\sum P(B \mid R \,\&\, \theta)P(\theta)$, where θ represents possible values of the percentage of ravens in the universe that

SCIENTIFIC REASONING: THE BAYESIAN APPROACH

COLIN HOWSON

AND

PETER URBACH

Second Edition

. . . if this [probability] calculus be condemned, then the whole of the sciences must also be condemned.
—Henri Poincaré

Our assent ought to be regulated by the grounds of probability.
—John Locke

Open Court
Chicago and La Salle, Illinois

1996