SCIENTIFIC REASONING
THE BAYESIAN APPROACH

Colin Howson and Peter Urbach

THIRD 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

To order books from Open Court, call toll-free 1-800-815-2280, or visit our website at www.opencourtbooks.com.

Open Court Publishing Company is a division of Carus Publishing Company.

First printing 2006

Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Howson, Colin.
Scientific reasoning : the Bayesian approach / Colin Howson and Peter Urbach.—3rd ed.
p. cm.
Includes bibliographical references (p.) and index.
ISBN-13: 978-0-8126-9578-6 (trade pbk. : alk. paper)
ISBN-10: 0-8126-9578-X (trade pbk. : alk. paper)
1. Science--Philosophy. 2. Reasoning. 3. Bayesian statistical decision theory.
I. Urbach, Peter. II. Title.
Q175.H87 2005
501—dc22

2005024868

CHAPTER 4
Bayesian Induction: Deterministic Theories

Philosophers of science have traditionally concentrated mainly on deterministic hypotheses, leaving statisticians to discuss how statistical, or non-deterministic theories should be assessed. Accordingly, a large part of what naturally belongs to philosophy of science is normally treated as a branch of statistics, going under the heading 'statistical inference'. It is not surprising therefore, that philosophers and statisticians have developed distinct methods for their different purposes. We shall follow the tradition of dealing separately with deterministic and statistical theories. As will become apparent however, we regard this separation as artificial and shall in due course explain how Bayesian principles provide a unified scientific method.

4.a Bayesian Confirmation

Information gathered in the course of observation is often considered to have a bearing on the merits of a theory or hypothesis (we use the terms interchangeably), either by confirming or disconfirming it. Such information may derive from casual observation or, more commonly, from experiments deliberately contrived with a view to obtaining relevant evidence. The idea that observations may count as evidence either for or against a theory, or be neutral towards it, is at the heart of scientific reasoning, and the Bayesian approach must start with a suitable understanding of these concepts.

As we have described, a very natural one is at hand, for if $P(h)$ measures your belief in a hypothesis when you do not know the evidence, and $P(h \mid e)$ is the corresponding measure when you do, e strengthens your belief in h or, we may say, confirms it, just in case the second probability exceeds the first. We refer in the usual way to $P(h)$ as 'the prior probability' of h, and to $P(h \mid e)$ as the 'posterior probability' of h relative to, or in the light of e, and we adopt the following definitions:

e **confirms or supports** h just in case $P(h \mid e) > P(h)$

e **disconfirms** h just in case $P(h \mid e) < P(h)$

e **is neutral towards** h just in case $P(h \mid e) = P(h)$.

One might reasonably take $P(h \mid e) - P(h)$ as measuring the *degree* of e's support for h, though other measures, involving for example the ratios of these terms, have also been suggested,[1] but disagreements on this score need not be settled in this book. We shall, however, say that when $P(h \mid e) > P(h \mid e') > P(h)$, the first piece of evidence confirms the hypothesis more than the second does.

According to Bayes's theorem, the posterior probability of a hypothesis depends on the three factors: $P(e \mid h)$, $P(e)$ and $P(h)$. Hence, if you know these, you can determine whether or not e confirms h, and more importantly, calculate $P(h \mid e)$. In practice, the various probabilities may be known only imprecisely, but as we shall show in due course, this does not undermine Bayes's theorem as a basis for scientific inference.

The dependence of the posterior probability on these three terms is reflected in three principal aspects of scientific inference. First, other things being equal, the more probable the evidence, relative to the hypothesis, the more that hypothesis is confirmed. At one extreme, if e refutes h, then $P(e \mid h) = 0$ and so disconfirmation is at a maximum, while the greatest confirmation is given when $P(e \mid h) = 1$, which will be met in practice when h logically implies e. Statistical hypotheses admit intermediate values for $P(e \mid h)$; as we show in later chapters, the higher the value, the greater the confirmation, other things being equal.

[1] For discussions of various other measures see, for example, Good 1950, and Jeffrey 2004, pp. 29–32.

Secondly, the power of e to confirm h depends on $P(e)$, that is, on the probability of e when h is not assumed to be true. This, of course, is not the same as the probability of e when h is assumed to be false; in fact $P(e)$ is related to the latter by the formula: $P(e) = P(e \mid h)P(h) + P(e \mid {\sim}h)P({\sim}h)$, as we showed in Chapter 2 (Theorem 12). This inverse dependence of $P(h \mid e)$ on $P(e)$ corresponds to the familiar intuition that the more surprising the evidence, the more confirmation it provides.

Thirdly, the posterior probability of a hypothesis depends on its prior probability, a dependence that is sometimes discernible in attitudes to so-called 'ad hoc' hypotheses and in the frequently expressed preference 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.

We shall, in the course of this chapter, examine each of these facets of inductive reasoning.

4.b Checking a Consequence

A characteristic pattern of scientific inference occurs when a logical consequence of a theory is shown to be false and the theory thereby refuted. As we saw, this sort of inference, with its unimpeachable logic, impressed Popper so much that he made it the centrepiece and guiding principle of his scientific philosophy. Bayesian philosophy readily accommodates the crucial features of a theory's refutation by empirical evidence. For if a hypothesis h entails a consequence e, then, as is easily shown, provided $P(h) > 0$, $P(e \mid h) = 1$ and $P(h \mid {\sim}e) = 0$. Interpreted in the Bayesian fashion, this means that h is maximally disconfirmed when it is refuted. Moreover, as we should expect, once a theory has been refuted, no further evidence can ever confirm it, unless the refuting evidence be revoked. For if e' is any other observation that is logically consistent with e, and if $P(h \mid {\sim}e)$ is zero, then so is $P(h \mid {\sim}e \,\&\, e')$.

Another characteristic pattern of scientific inference occurs when a logical consequence of a theory is shown to be true and the theory then regarded as confirmed. Bayes's theorem shows why and under what circumstances a theory is confirmed by its consequences. First, it follows from the theorem that a theory is always confirmed by a logical consequence, provided neither the evidence nor the theory takes either of the extreme probability values. For if h entails e, $P(e \mid h) = 1$, so that $P(h \mid e) = \frac{P(h)}{P(e)}$. Hence, provided $0 < P(e) < 1$ and $P(h) > 0$, $P(h \mid e) > P(h)$, which means that e confirms h.

Secondly, the probability axioms tell us, correctly, that succeeding confirmations by logical consequences eventually diminish in force (Jeffreys 1961, pp. 43–44). For let $e_1, e_2, \ldots, e_n$ be a succession of logical consequences of h, then

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

As we showed earlier, if h entails all the e_i, then $P(h \mid e_1 \,\&\, \ldots \,\&\, e_n) \geq P(h \mid e_1 \,\&\, \ldots \,\&\, e_{n-1})$. It follows from the Bolzano-Weierstrass theorem that the non-decreasing sequence of posterior probabilities has a limit. Clearly, the limits, as n tends to infinity, of the two posterior probabilities in this equation are the same, *viz*, $\lim P(h \mid e_1 \,\&\, \ldots \,\&\, e_n) = \lim P(h \mid e_1 \,\&\, \ldots \,\&\, e_{n-1})$. Hence, provided that $P(h) > 0$, $P(e_n \mid e_1 \,\&\, \ldots \,\&\, e_{n-1})$ must tend to 1. This explains why it is not sensible to test a hypothesis indefinitely. The result does not however tell us the precise point beyond which further predictions of the hypothesis are sufficiently probable not to be worth examining, for that would require a knowledge of individuals' belief structures which logic does not supply.

A third salient feature of confirmation by a theory's consequences is that in many instances, specific categories of those consequences each have their own, limited capacity to confirm. This is an aspect of the familiar phenomenon that however often a particular experiment is repeated, its results can confirm a general theory only to a limited extent; and when an experiment's capacity to generate significant confirming evidence for the theory has been exhausted through repetition, further support is sought from other experiments, whose outcomes are predicted by other parts of the theory.[2]

This phenomenon has a Bayesian explanation (Urbach 1981). Consider a general hypothesis h and let h_r be a substantial restriction of that hypothesis. A substantial restriction of Newton's theory might, for example, express the idea that freely falling bodies near the Earth's surface descend with constant acceleration, or that the period and length of a pendulum are related by the familiar formula. Since h entails h_r, $P(h) \leq P(h_r)$, as we showed in Chapter 2, and if h_r is much less speculative than its progenitor, it will often be much more probable.

Now consider a series of predictions that are implied by h, and 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 yields

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

Since the maximum value of the last probability in this equation is 1, it follows that however many predictions of h_r have been verified, the posterior probability of the main theory, h, can never rise above $\frac{P(h)}{P(h_r)}$. Therefore, the prior probability of h_r determines a limit to how far evidence entailed by it can confirm h. And this explains the phenomenon under consideration, for the predictions verified by means of an experiment (that is, a procedure designed

[2] This is related to the phenomenon that the more varied a body of evidence, the greater its inductive force, which we discuss in section 4.g below.

to a specified pattern) do normally follow from and confirm a much-restricted version of the predicting theory.

The arguments and explanations in this section rely on the possibility that evidence already accumulated from an experiment can increase the probability that further performances of the experiment will produce similar results. Such a possibility was denied by Popper and by his supporters, on the grounds that the probabilities involved are not objective. How then do they explain the fact, familiar to every scientist, that repeating some experiment indefinitely (or usually, more than a very few times) is pointless? Musgrave (1975) attempted an explanation. He argued that after a certain (unspecified) number of repetitions, the scientist should form a generalization to the effect that the experiment will always yield a result that is similar, in certain respects to those results already obtained, and that this generalization should then be entered into 'background knowledge'. Relative to the newly augmented background knowledge, the experiment is certain to produce the same result when it is next performed as it did before. Musgrave then appealed to the putative principle, which we discuss in the next section, that evidence confirms a hypothesis in proportion to the difference between its probability relative to the hypothesis plus background knowledge and its probability relative to background knowledge alone, that is, to $P(e \mid h \,\&\, b) - P(e \mid b)$, and inferred that even if the experiment did produce the expected result when next conducted, the hypothesis would receive no new confirmation.

A number of decisive objections can be raised against this account. First, as we show in the next section, although it forms part of the Bayesian account and seems to be a feature of science that confirmation depends in its degree upon the probability of the evidence, that principle has no basis in Popperian methodology. Popper simply invoked it ad hoc. Secondly, Musgrave's suggestion takes no account of the fact that particular experimental results may be generalized in infinitely many ways. This is a substantial objection since different generalizations give rise to different implications about future experimental outcomes. So Musgrave's explanation calls for a rule that would guide the scientist to a particular and appropriate generalization; but we cannot see how appropriateness could be defined or such a rule possibly justified within the limitations of Popperian philosophy. Finally, the decision to designate the generalization background *knowledge*, with the effect that has on our evaluation of other theories and on our future conduct regarding for example, whether or not to repeat certain experiments, is comprehensible only if we have invested some confidence in the generalization. But then this Popperian account tacitly invokes the same kind of inductive notion as it was designed to avoid. The fact is that the phenomena concerning the confirming power of experiments and their repetitions are essentially inductive and are beyond the reach of anti-inductivist methodologies such as Popper's.

4.c The Probability of the Evidence

In the Bayesian account, confirmation occurs when the posterior probability of a hypothesis exceeds its prior probability, and the greater the difference, the greater the confirmation. Now Bayes's theorem may be expressed in the following ways:

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

We see that the evidential force of e is entirely expressed by the ratio $\dfrac{P(e \mid \sim h)}{P(e \mid h)}$, known as the *Bayes factor*. The smaller this factor, that is to say, the more probable the evidence if the hypothesis is true than if it is false, the greater is the confirmation. In the deterministic case, where h entails e, so that $P(e \mid h) = 1$, confirmation depends inversely on $P(e)$ or $P(e \mid \sim h)$; this fact is 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 some time and you do, your faith in his powers of precognition would not be much enhanced: you would probably continue to regard his predictions as simply guesswork. But if the prediction also gave you 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 nicely with an incident in Shakespeare's *Macbeth*. The three witches, using their special brand of divination, tell Macbeth that he will soon become both Thane of Cawdor and King of Scotland. Macbeth finds these two predictions incredible:

> 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 shortly after making this declaration, he learns that the Thane of Cawdor prospered no longer, was in fact condemned to death, and that he, Macbeth, had succeeded to the title, whereupon, his attitude to the witches' powers of foresight alters entirely, and he comes to believe their other predictions.

Charles Babbage (1827), the celebrated polymath and 'father of computing', examined numerous logarithmic tables published over two centuries in various parts of the world, with a view to determining whether they derived from a common source or had been worked out independently. He found the same six errors in all but two and drew the "irresistible" conclusion that the tables containing those errors had been copied from a single original. As Jevons (1874, pp. 278–79) pointed out, the force of this conclusion springs from the fact that if the tables originated from the same source, then it is practically certain that an error in one will be reproduced in the others, but if they did not, the probability of errors being duplicated is minuscule. Such reasoning is so compelling that compilers of mathematical tables regularly protect their copyrights by purposely incorporating some minor errors "as a trap for would-be plagiarists" (L.J. Comrie)[3]; and cartographers do the same.

The inverse relationship between the probability of evidence and its confirming power is a simple and direct consequence of

---

[3] This is quoted in Bowden 1953, p. 4.

Bayesian theory. On the other hand, methodologies that eschew probabilistic evaluations of hypotheses, in the interests of objectivity, seem constitutionally unable to account for the phenomenon. Popper (1959a, appendix *ix) recognized the need to provide such an account and rose to the challenge. First, he conceded that, in regard to confirmation, the significant quantities are $P(e \mid h)$ and $P(e)$; he then measured the amount of confirmation or "corroboration" which e confers on h by the difference between those quantities. But Popper never said explicitly what he meant by the probability of evidence. He could not allow it a subjective connotation without compromising the intended objectivist quality of his methodology, yet he never worked out what objective significance the term could have. His writings suggest he had in mind some purely logical notion of probability, but neither he nor anyone else has managed to give an adequate account of logical probability. Secondly, Popper never satisfactorily justified his claim that hypotheses benefit in any epistemic sense from improbable evidence; indeed, the idea has been closely examined by philosophers and is generally regarded as indefensible within the Popperian scheme. (See Chapter 1, above, and, for example, Howson 2000, and 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 9, where some other criticisms of the Bayesian approach are rebutted.

4.d The Ravens Paradox

The Bayesian position that confirmation is a matter of degree, determined by Bayes's theorem, scotches a famous puzzle, first posed by Hempel (1945), known as the *Paradox of Confirmation* or sometimes as the *Ravens Paradox*. It was called a paradox because its premises seemed extremely plausible, despite their supposedly counter-intuitive consequences, and the reference to ravens stems from the paradigm hypothesis, 'All ravens are black', that is frequently used to present the problem. The alleged difficulty arises from the following assumptions about confirmation. (RB will signify the proposition that a certain object is black and a raven, and $\bar{R}\bar{B}$ that it is neither black nor a raven.)

1. Hypotheses of the form 'All Rs are B' are confirmed by evidence of something that is both R and B. (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.*)

Now, by the Nicod Condition, 'All non-Bs are non-Rs' is confirmed by $\bar{R}\bar{B}$; and by the Equivalence Condition, so is 'All Rs are B', since the two generalizations are logically equivalent. Many philosophers regard this consequence as blatantly false, since it says that you can confirm the hypothesis that all ravens are black by observing a non-black non-raven, say, a white lie or a red herring. This seems to suggest that you could investigate that and other similar generalizations *just as well* by examining objects on your desk as by studying ravens on the wing. But that would be a non sequitur. For the fact that RB and $\bar{R}\bar{B}$ both confirm a hypothesis does not mean that they do so with equal force. And once it is recognized that confirmation is a matter of degree, the conclusion ceases to be counter-intuitive, because it is compatible with $\bar{R}\bar{B}$ confirming 'All Rs are B', but to a negligible degree. This simple point constitutes the Bayesian solution to the problem.

But a Bayesian analysis can take the matter further, first of all, by demonstrating that in the case of the paradigm hypothesis, data of the form $\bar{R}\bar{B}$ do in fact confirm to a negligible degree; secondly, by showing that Nicod's condition is not valid as a universal principle of confirmation. Consider the first point. The impact of the two data on h, 'All ravens are black', is given as follows:

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

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 close approximation, though Horwich (1982, p. 59) thinks it lacks plausibility.[4] By parallel 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 $P(B \mid R) = \Sigma P(B \mid R \,\&\, \theta)P(\theta \mid R) = \Sigma P(B \mid R \,\&\, \theta)P(\theta)$, where θ represents possible values of the proportion of ravens that are black (h says that $\theta = 1$), and assuming independence between θ and R. Finally, $P(B \mid R \,\&\, \theta) = \theta$, for if the proportion of ravens in the universe that are black is θ, the probability of a randomly selected raven being black is also θ.

Combining all these considerations with Bayes's theorem yields:

$$\frac{P(h \mid RB)}{P(h)} = \frac{1}{\Sigma\theta P(\theta)} \;\&\; \frac{P(h \mid \bar{R}\bar{B})}{P(h)} = \frac{1}{P(\bar{R} \mid \bar{B})}.$$

According to the first of these equations, the ratio of the posterior to the prior probabilities of h is inversely proportional to $\Sigma\theta P(\theta)$. This means, for example, that if it were initially very probable that all or virtually all ravens are black, then $\Sigma\theta P(\theta)$ would be large and RB would confirm h rather little. While if it were initially relatively probable that most ravens are not black, the confirmation could be substantial. Intermediate degrees of uncertainty regarding θ would bring their own levels of confirmation to h.

The second equation refers to the confirmation to be derived from the observation of a non-black non-raven, and here the crucial probability term is $P(\bar{R} \mid \bar{B})$. Now presumably there are vastly more non-black things in the universe than ravens. So even if we felt certain that no ravens are black, the probability of some object about which we know nothing, except that it is not black, being a non-raven must be very high, practically 1. Hence, $P(h \mid \bar{R}\bar{B}) = (1 - \varepsilon)P(h)$, where ε is a very small positive number;

---

[4] Vranas (2004) interprets the assumption as asserting that whether some arbitrary object is a raven "should" be independent of h, and he criticizes this and other Bayesian accounts for depending upon a claim for which, he says, there can be no reasoned defence. But our argument does not need such a strong assumption. Our position is merely that in this particular case, our and, we suspect, most other people's personal probabilities are such that independence obtains.

therefore, observing that some object is neither a raven nor black provides correspondingly little confirmation for *h*.[5]

A Bayesian analysis necessarily retains the Equivalence Condition but gives only qualified backing to the Nicod Condition, for it anticipates circumstances in which the condition fails. For instance, suppose the hypothesis under examination is 'All grasshoppers are located outside the County of Yorkshire'. One of these creatures appearing just beyond the county border is an instance of the generalization and, according to Nicod, confirms it. But it might be more reasonably argued that since there are no border controls or other obstacles to restrict the movement of grasshoppers in that area, the observation of one on the edge of the county but outside it increases the probability that some others have actually crossed over, and hence, contrary to Nicod, it undermines the hypothesis. In Bayesian terms, this is a case where, relative to background information, the probability of some datum is reduced by a hypothesis—that is, $P(e \mid h) < P(e)$—which is thereby disconfirmed—that is, $P(h \mid e) < P(h)$.[6] This example is adapted from Swinburne 1971, though the idea seems to originate with Good 1961.

Another, more striking case where Nicod's Condition breaks down was invented by Rosenkrantz (1977, p. 35). Three people leave a party, each with a hat. The hypothesis that none of the three has his own hat is confirmed, according to Nicod, by the observation that person 1 has person 2's hat and by the observation that person 2 has person 1's hat. But since the hypothesis concerns only three, particular people, the second observation must *refute* the hypothesis, not confirm it.

Our grasshopper example may also be used to show that instances of the type $\bar{R}B$ can sometimes confirm 'All *R*s are *B*'. Imagine that an object that looks for all the world like a grasshopper had been found hopping about just outside Yorkshire and that it turned out to be some other sort of insect. The discovery that the object was not a grasshopper after all would be relatively unlikely unless the grasshopper hypothesis were true (hence, $P(e) < P(e \mid h)$);

[5] The account given here is substantially similar to Mackie's, 1963.

[6] This example is adapted from Swinburne 1971, though the idea seems to originate with Good 1961.

so it would confirm that hypothesis. If the deceptively grasshopper-like object were discovered within the county, the same conclusion would follow for this $\bar{R}\bar{B}$ instance.

Horwich (1982, p. 58) has argued that the ravens hypothesis could be differently confirmed depending on how the black raven was chosen, either by randomly selecting an object from the population of ravens, or by restricting the selection to the population of black objects. Korb (1994) provides a convincing demonstration of this, which we discuss in a related context in Chapter 8.

We do not accept Horwich's argument for his conclusion. Denoting a black raven either R^*B or RB^*, depending on whether it was discovered by the first selection process or the second, he claims that evidence of the former kind always confirms more, because only it subjects the raven hypothesis to the risk of falsification. But this conflates the process of collecting evidence, which may indeed expose the hypothesis to different risks of refutation, with the evidence itself, which either does or does not refute the hypothesis, and in the present case it does not.

Our conclusions are, first, that the so-called paradox of the ravens is not in fact problematic; secondly, that of the two conditions of confirmation that generated it only the Equivalence Condition is acceptable; and thirdly, that Bayesian theory explains why.

4.e The Duhem Problem

The Duhem (sometimes called the Duhem-Quine) problem arises with philosophies of science of the type associated with Popper, which emphasize the power of certain evidence to refute a theory. According to Popper, falsifiability is the feature of a theory which makes it scientific. "Statements or systems of statements," he said, "in order to be ranked as scientific, must be capable of conflicting with possible, or conceivable, observations" (1963, p. 39). And claiming to apply this criterion, he judged Einstein's gravitational theory scientific and Freud's psychology not. The term 'scientific' carries a strong flavour of commendation, which is, however, misleading in this context. For Popper could never demonstrate a link between his concept of scientificness and epistemic or inductive merit: a theory that is scientific in Popper's sense is not necessarily true, or probably true, nor can it be said either definitely or even probably to lead to the truth. There is little alternative then, in our judgment, to regarding Popper's demarcation between scientific and unscientific statements as without normative significance, but as a claim about the content and character of what is ordinarily termed science.

Yet as an attempt to understand the practice of science, Popper's ideas bear little fruit. First of all, the claim that scientific theories are falsifiable by "possible, or conceivable, observations" raises a difficulty, because an observation can only falsify a theory (in other words conclusively demonstrate its falsity) if it is itself conclusively certain. Yet as Popper himself appreciated, no observations fall into this category; they are all fallible. But unwilling to concede degrees of fallibility or anything of the kind, Popper took the view that observation reports that are admitted as evidence "are accepted as the result of a decision or agreement; and to that extent they are *conventions*" (1959a, p. 106; our italics). It is unclear to what psychological attitude such acceptance corresponds, but whatever it is, Popper's view pulls the rug from under his own philosophy, since it implies that no theory can really be falsified by evidence. Every 'falsification' is merely a convention or decision: "From a logical point of view, the testing of a theory depends upon basic statements whose acceptance or rejection, in its turn, depends upon our *decisions*. Thus it is *decisions* which settle the fate of theories" (1959a, p. 108).

Watkins was one of those who saw that the Popperian position could not rest on this arbitrary basis, and he attempted to shore it up by arguing that some infallibly true observation statements do in fact exist. He agreed that a statement like 'the hand on this dial is pointing to the numeral 6' is fallible, since it is possible, however unlikely, that the person reporting the observation mistook the position of the hand. But he claimed that introspective perceptual reports, such as 'in my visual field there is now a silvery crescent against a dark blue background', "may rightly be regarded by their authors when they make them as infallibly true" (1984, pp. 79 and 248). But in our opinion Watkins was wrong, and the statements he regarded as infallible are open to the same sceptical doubts as any other observational report. We can illustrate this through the above example: clearly it is possible, though admittedly not very probable, that the introspector has misremembered and mistaken the shape he usually describes as a crescent, or the sensation he usually receives on reporting blue and silvery images. These and other similar sources of error ensure that introspective reports are not exempt from the rule that non-analytic statements are fallible.

Of course, the kinds of observation statement we have mentioned, if asserted under appropriate circumstances, would never be seriously doubted, for although they could be false, they have a force and immediacy that carries conviction: in the traditional phrase, they are 'morally certain'. But if they are merely indubitable, then whether or not a theory is regarded as refuted by observational data rests ultimately on a subjective feeling of certainty, a fact that punctures the objectivist pretensions of Popperian philosophy.

A second objection to Popper's falsifiability criterion, and the one upon which we shall focus for its more general interest, is that it deems unscientific most of those theories that are usually judged science's greatest achievements. This is the chief aspect of the well-known criticisms advanced by Polanyi (1962), Kuhn (1970), and Lakatos (1970), amongst others, but based on the arguments of Duhem (1905). They pointed out that notable theories of science are typically unfalsifiable by observation statements, because they only make empirical predictions in association with certain auxiliary theories. Should any such prediction turn out to be false, logic does not compel us to regard the principal theory as untrue, since the error may lie in one or more of the auxiliaries. Indeed, there are many occasions in the history of science when an important theory led to a false prediction but was not itself significantly impugned thereby. The problem that Duhem posed was this: *when several distinct theories are involved in deriving a false prediction, which of them should be regarded as false?*

Lakatos and Kuhn on the Duhem Problem

Lakatos and Kuhn both investigated scientific responses to anomalies and were impressed by the tendency they observed for the benefit of the doubt persistently to be given to particular, especially fundamental theories, and for one or more of the auxiliary theories regularly to be blamed for any false prediction. Lakatos drew from this observation the lesson that science of the most significant kind usually proceeds in what he called scientific research programmes, each comprising a central, or 'hard core', theory, and a so-called 'protective belt' of auxiliary theories. During the lifetime of a research programme, these elements are combined to yield empirical predictions, which are then experimentally checked; and if they turn out to be false, the auxiliary hypotheses act as a protective shield, as it were, for the hard core, and take the brunt of the refutation. A research programme is also characterised by a set of heuristic rules by which it develops new auxiliary hypotheses and extends into new areas. Lakatos regarded Newtonian physics as an example of a research programme, the three laws of mechanics and the law of gravitation comprising the hard core, and various optical theories, propositions about the natures and dispositions of the planets, and so forth, being the protective belt.

Kuhn's theory is similar to the methodology we have just outlined and probably inspired it in part. Broadly speaking, Kuhn's 'paradigm' is the equivalent of a scientific research programme, though his idea is developed in less detail.

Lakatos, following Popper, also added a normative element, something that Kuhn deliberately avoided. He held that it was perfectly all right to treat the hard core systematically as the innocent party in a refutation, provided the research programme occasionally leads to successful "novel" predictions or to successful, "non-ad hoc" explanations of existing data. Lakatos called such programmes "progressive."

> The sophisticated falsificationist [which Lakatos counted himself] . . . sees nothing wrong with a group of brilliant scientists conspiring to pack everything they can into their favourite research programme . . . with a sacred hard core. As long as their genius—and luck—enables them to expand their programme '*progressively*', while sticking to its hard core, they are allowed to do it. (Lakatos 1970, p. 187)

If, on the other hand, the research programme persistently produces false predictions, or if its explanations are habitually ad hoc, Lakatos called it "degenerating." The notion of an ad hoc explanation—briefly, one that does not produce new and verified predictions—is central to attempts by the Popperian school to deal with the Duhem problem and we discuss it in greater detail below, in section **g**. In appraising research programmes, Lakatos employed the tendentious terms 'progressive' and 'degenerating', but he never succeeded in substantiating their normative intimations, and in the end he seems to have abandoned the attempt and settled on the more modest claim that, as a matter of historical fact, progressive programmes were well regarded by scientists, while degenerating ones were distrusted and eventually dropped.

This last claim, it seems to us, contains a measure of truth, as evidenced by case studies in the history of science, such as those in Howson 1976. But although Lakatos and Kuhn identified and described an important aspect of scientific work, they could not explain it or rationalize it. So, for example, Lakatos did not say why a research programme's occasional predictive success could compensate for numerous failures, nor did he specify how many such successes are needed to convert a degenerating programme into a progressive one, beyond remarking that they should occur "now and then".

Lakatos was also unable to explain why certain theories are raised to the privileged status of hard core in a research programme while others are left to their own devices. His writings give the impression that the scientist is free to decide the question at will, by "methodological fiat", as he says. Which suggests that it is perfectly canonical scientific practice to set up any theory whatever as the hard core of a research programme, or as the central pattern of a paradigm, and to attribute all empirical difficulties to auxiliary hypotheses. This is far from being the case. For these reasons and also because of difficulties with the notion of an ad hoc hypothesis, to be discussed below, neither Kuhn's theory of paradigms nor Lakatos's so-called 'sophisticated falsificationism' are in any position to solve the Duhem problem.

The Bayesian Resolution

The questions left unanswered in the Kuhn and Lakatos methodologies are addressed and resolved, as Dorling (1979) brilliantly showed, by referring to Bayes's theorem and considering how the individual probabilities of theories are severally altered when, as a group, they have been falsified.

We shall illustrate the argument through a historical example that Lakatos (1970, pp. 138–140; 1968, pp. 174–75) drew heavily upon. In the early nineteenth century, William Prout (1815, 1816), a medical practitioner and chemist, advanced the idea that the atomic weight of every element is a whole-number multiple of the atomic weight of hydrogen, the underlying assumption being that all matter is built up from different combinations of some basic element. Prout believed hydrogen to be that fundamental building block. Now many of the atomic weights recorded at the time were in fact more or less integral multiples of the atomic weight of hydrogen, but some deviated markedly from Prout's expectations. Yet this did not shake the strong belief he had in his hypothesis, for in such cases he blamed the methods that had been used to measure those atomic weights. Indeed, he went so far as to adjust the atomic weight of the element chlorine, relative to that of hydrogen, from the value 35.83, obtained by experiment, to 36, the nearest whole number. Thomas Thomson (1818, p. 340) responded in a similar manner when confronted with 0.829 as the measured atomic weight (relative to the atomic weight of oxygen) of the element boron, changing it to 0.875, "because it is a multiple of 0.125, which all the atoms seem to be". (Thomson erroneously took the relative atomic weights of hydrogen and oxygen as 0.125.)

Prout's reasoning relative to chlorine and Thomson's, relative to boron, can be understood in Bayesian terms as follows: Prout's hypothesis t, together with an appropriate assumption a, asserting the accuracy (within specified limits) of the measuring techniques, the purity of the chemicals employed, and so forth, implies that the ratio of the measured atomic weights of chlorine and hydrogen will approximate (to a specified degree) a whole number. In 1815 that ratio was reported as 35.83—call this the evidence e—a value judged to be incompatible with the conjunction of t and a.

The posterior and prior probabilities of t and of a are related by Bayes's theorem, as follows:

$$P(t \mid e) = \frac{P(e \mid t)P(t)}{P(e)} \text{ and } P(a \mid e) = \frac{P(e \mid a)P(a)}{P(e)} .$$

To evaluate the two posterior probabilities, it is necessary to quantify the various terms on the right-hand sides of these equations.

Consider first the prior probabilities of t and of a. J.S. Stas, a distinguished Belgian chemist whose careful atomic weight measurements were highly influential, gives us reason to think that chemists of the period were firmly disposed to believe in t, recalling that "In England the hypothesis of Dr Prout was almost universally accepted as absolute truth" and that when he started investigating the subject, he himself had "had an almost absolute confidence in the exactness of Prout's principle" (1860, pp. 42 and 44).

It is less easy to ascertain how confident Prout and his contemporaries were in the methods used to measure atomic weights, but their confidence was probably not great, in view of the many clear sources of error. For instance, errors were recognised to be inherent in the careful weighings and manipulations that were required; the particular chemicals involved in the experiments to measure the atomic weights were of questionable purity; and, in those pioneer days, the structures of chemicals were rarely known with certainty.[7] These various uncertainties were reinforced by the fact that independent measurements of atomic weights, based on the transformations of different chemicals, rarely delivered identical results.[8] On the other hand, the chemists of the time must have felt that that their atomic weight measurements were more likely to be accurate than not, otherwise they would hardly have reported them.[9]

---

[7] The several sources of error were rehearsed by Mallet (1893).

[8] For example, Thomson (1818, p. 340) reported two independent measurements—2.998 and 2.66—for the weight, relative to the atomic weight of oxygen, of a molecule of boracic (boric) acid. He required this value in order to calculate the atomic weight of boron from the weight of the boric acid produced after the element was combusted.

[9] "I am far from flattering myself that the numbers which I shall give are all accurate; on the contrary, I have not the least doubt that many of them are still erroneous. But they constitute at least a nearer approximation to the truth than the numbers contained in the first table [which Thomson had published some years before]" (Thomson 1818, p. 339).

For these reasons, we conjecture that $P(a)$ was in the neighbourhood of 0.6 and that $P(t)$ was around 0.9, and these are the figures we shall work with. We stress that these figures and those we shall assign to other probabilities are intended chiefly to show that hypotheses that are jointly refuted by an observation, may sometimes be disconfirmed to very different degrees, so illustrating the Bayesian resolution of Duhem's problem. Nevertheless, we believe that the figures we have suggested are reasonably accurate and sufficiently so to throw light on the historical progress of Prout's hypothesis. As will become apparent, the results we obtain are not very sensitive to variations in the assumed prior probabilities.

The posterior probabilities of t and of a depend also on $P(e)$, $P(e \mid t)$, and $P(e \mid a)$. Using the theorem of total probability, the first two of these terms can be expressed as follows:

$$P(e) \quad = P(e \mid t)P(t) + P(e \mid \sim t)P(\sim t)$$

$$P(e \mid t) = P(e \mid t \,\&\, a)P(a \mid t) + P(e \mid t \,\&\, \sim a)P(\sim a \mid t).$$

We will follow Dorling in taking t and a to be independent, viz, $P(a \mid t) = P(a)$ and hence, $P(\sim a \mid t) = P(\sim a)$. As Dorling points out (1996), this independence assumption makes the calculations simpler but is not crucial to the argument. Nevertheless, that assumption accords with many historical cases and seems clearly right here. For we put ourselves in the place of chemists of Prout's day and consider how our confidence in his hypothesis would have been affected by a knowledge that particular chemical samples were pure, that particular substances had particular molecular structures, specific gravities, and so on. It seems to us that it would not be affected at all. Bovens and Hartmann (2003, p. 111) take a different view and have objected to the assumption of independence in this context. Speaking in general terms, they allege that "experimental results are determined by a hypothesis and auxiliary theories that are often hopelessly interconnected with each other."

> And these interconnections raise havoc in assessing the value of experimental results in testing hypotheses. There is always the fear that the hypothesis and the auxiliary theory really come out of the

same deceitful family and that the lies of one reinforce the lies of the other.

We do not assert that theories are never entangled in the way that Bovens and Hartmann describe, but for the reasons we have just cited, it seems to us that the present situation is very far from being a case in point.

Returning to the last equation, if we incorporate the independence assumption and take account of the fact that since the conjunction t & a is refuted by e, $P(e \mid t \& a)$ must be zero, we obtain:

$$P(e \mid t) = P(e \mid t \& {\sim}a)P({\sim}a).$$

By parallel reasoning, we may derive the results:

$$P(e \mid a) = P(e \mid {\sim}t \& a)P({\sim}t)$$

$$P(e \mid {\sim}t) = P(e \mid {\sim}t \& a)P(a) + P(e \mid {\sim}t \& {\sim}a)P({\sim}a).$$

So, provided the following terms are fixed, which we have done in a tentative way, to be justified presently, the posterior probabilities of t and of a can be calculated:

$$P(e \mid {\sim}t \& a) = 0.01$$

$$P(e \mid {\sim}t \& {\sim}a) = 0.01$$

$$P(e \mid t \& {\sim}a) = 0.02.$$

The first of these gives the probability of the evidence if Prout's hypothesis is not true, but if the assumptions made in calculating the atomic weight of chlorine are accurate. Certain nineteenth-century chemists thought carefully about such probabilities, and typically took a theory of random distribution of atomic weights as the alternative to Prout's hypothesis (for instance, Mallet 1880); we shall follow this. Suppose it had been established for certain that the atomic weight of chlorine lies between 35 and 36. (The final results we obtain respecting the posterior probabilities of t and of a are, incidentally, unaffected by the width of this interval.) The random-distribution theory assigns equal probabilities to the atomic weight of an element lying in any 0.01-wide band. Hence, on the assumption that a is true, but t false, the probability that the atomic weight of chlorine lies in the interval 35.825 to 35.835 is 0.01. We have attributed the same value to $P(e \mid \sim t \ \& \sim a)$, on the grounds that if a were false, because, say, some of the chemicals were impure, or had been inaccurately weighed, then, still assuming t to be false, one would not expect atomic weights to be biased towards any particular part of the interval between adjacent integers.

We have set the probability $P(e \mid t \ \& \sim a)$ rather higher, at 0.02. The reason for this is that although some impurities in the chemicals and some degree of inaccuracy in the measurements were moderately likely at the time, chemists would not have considered their techniques entirely haphazard. Thus if Prout's hypothesis were true and the measurement technique imperfect, the measured atomic weights would be likely to deviate somewhat from integral values; but the greater the deviation, the less the likelihood, so the probability distribution of atomic weight measurements falling within the 35–36 interval would not be uniform, but would be more concentrated around the whole numbers.

Let us proceed with the figures we have proposed for the crucial probabilities. We note however that the absolute values of the probabilities are unimportant, for, in fact, only their relative values count in the calculation. Thus we would arrive at the same results with the weaker assumptions that $P(e \mid \sim t \ \& \ a) = P(e \mid \sim t \ \& \sim a) = \frac{1}{2}P(e \mid t \ \& \sim a)$. We now obtain:

$$P(e \mid \sim t) = 0.01 \times 0.6 + 0.01 \times 0.4 = 0.01$$

$$P(e \mid t) = 0.02 \times 0.4 = 0.008$$

$$P(e \mid a) = 0.01 \times 0.1 = 0.001$$

$$P(e) = 0.008 \times 0.9 + 0.01 \times 0.1 = 0.0082.$$

Finally, Bayes's theorem allows us to derive the posterior probabilities in which we are interested:

$P(t \mid e) = 0.878$ (Recall that $P(t) = 0.9$)

$P(a \mid e) = 0.073$ (Recall that $P(a) = 0.6$).

We see then that the evidence provided by the measured atomic weight of chlorine affects Prout's hypothesis and the set of auxiliary hypotheses very differently; for while the probability of the first is scarcely changed, that of the second is reduced to a point where it has lost all credibility.

It is true that these results depend upon certain—we have argued plausible—premises concerning initial probabilities, but this does not seriously limit their general significance, because quite substantial variations in the assumed probabilities lead to quite similar conclusions, as the reader can verify. So for example, if the prior probability of Prout's hypothesis were 0.7 rather than 0.9, the other assignments remaining unchanged, $P(t \mid e)$ would equal 0.65, and $P(a \mid e)$ would be 0.21. Thus, as before, Prout's hypothesis is still more likely to be true than false in the light of the adverse evidence, and the auxiliary assumptions are still much more likely to be false than true.

Successive pieces of adverse evidence may, however, erode the probability of a hypothesis so that eventually it becomes more likely to be false than true and loses its high scientific status. Such a process would correspond to a Lakatosian degenerating research programme or be the prelude to a Kuhnian paradigm shift. In the present case, the atomic weight of chlorine having been repeated in various, improved ways by Stas, whose laboratory skill was universally recognized, Mallet (1893, p. 45) concluded that "It may be reasonably said that probability is against the idea of any future discovery . . . ever making the value of this element agree with an integer multiple of the atomic weight of hydrogen". And in the light of this and other atomic weight measurements he regarded Prout's original idea as having been "shown by the calculus of probability to be a very improbable one". And Stas himself, who started out so very sure of its truth, reported in 1860 that he had now "reached the complete conviction, the entire certainty, as far as certainty can be attained on such a subject, that Prout's law . . . is nothing but an illusion" (1860, p. 45).

We conclude that Bayes's theorem provides a framework that resolves the Duhem problem, unlike the various non-probabilistic methodologies which philosophers have sought to apply to it. And the example of Prout's hypothesis, as well as others that Dorling

(1979 and 1996) has analysed, show in our view, that the Bayesian model is essentially correct.

4.f Good Data, Bad Data, and Data Too Good to Be True

Good Data

The marginal influence that an anomalous observation may exert on a theory's probability contrasts with the dramatic effect of some confirmations. For instance, if the measured atomic weight of chlorine had been a whole number, in line with Prout's hypothesis, so that $P(e \mid t \,\&\, a) = 1$ instead of 0, and if the other probability assignments remained the same, the probability of the hypothesis would shoot up from a prior of 0.9 to a posterior of 0.998. And even more striking: had the prior probability of t been 0.7, its posterior probability would have risen to 0.99.

This asymmetry between the effects of anomalous and confirming instances was emphasized by Lakatos, who regarded it as highly significant in science, and as a characteristic feature of a research programme. He maintained that a scientist involved in such a programme typically "forges ahead with almost complete disregard of 'refutations'," provided there are occasional predictive successes (1970, p. 137): the scientist is "encouraged by nature's YES, but not discouraged by its NO" (p. 135). As we have indicated, we believe there to be much truth in Lakatos's observations; the trouble, however, is that these observations are merely absorbed, without justification, into his methodology; the Bayesian methodology, on the other hand, explains why and under what circumstances the asymmetry effect is present.

Bad Data

An interesting fact that emerges from the Bayesian analysis is that a successful prediction derived from a combination of two theories does not necessarily redound to the credit of both of them, indeed one may even be discredited. Consider Prout's hypothesis again, and suppose the atomic weight of chlorine had been determined, not in the established way, but by concentrating hard on the element while selecting a number blindly from a given range of numbers. And let us suppose that the atomic weight of chlorine is reported by this method to be a whole number. This is just what one would predict on the basis of Prout's hypothesis, if the outlandish measuring technique were accurate. But accuracy is obviously most unlikely, and it is equally obvious that the results of the technique could add little or nothing to the credibility of Prout's hypothesis. This intuition is upheld by Bayes's theorem: as before, let t be Prout's hypothesis and a the assumption that the measuring technique is accurate. Then, set $P(e \mid t \,\&\, {\sim}a) = P(e \mid {\sim}t \,\&\, {\sim}a) = P(e \mid {\sim}t \,\&\, a) = 0.01$, for reasons similar to those stated above. And, because, as we said, a is extremely implausible, we will set $P(a)$ at, say 0.0001. It then follows that t is not significantly confirmed by e, for $P(t)$ and $P(t \mid e)$ are virtually identical.

This example shows that Leibniz was wrong to declare as a maxim that "It is the greatest commendation of a hypothesis (next to truth) if by its help predictions can be made even about phenomena or experiments not [yet] tried". Leibniz, and Lakatos, who quoted these words with approval (1970, p. 123), seem to have overlooked the fact that if a prediction can be deduced from a hypothesis only with the assistance of highly questionable auxiliary claims, then that hypothesis may accrue very little credit when the prediction is verified. This explains why the various sensational predictions that Velikovsky drew from his theory failed to impress most serious astronomers, even when some of those predictions were to their amazement fulfilled. For instance, Velikovsky's prediction (1950, p. 351) of the existence of large quantities of petroleum on the planet Venus relied not only on his pet theory that various natural disasters in the past had been caused by collisions between the Earth and a comet, but also on a string of unsupported and implausible assumptions, for instance, that the comet in question carried hydrogen and carbon; that these had been converted to petroleum by electrical discharges supposedly generated in the violent impact with the Earth; that the comet had later evolved into the planet Venus; and some others. (More details of Velikovsky's theory are given in the next section.)

Data Too Good to Be True

Data are sometimes said to be 'too good to be true', when they fit a favoured hypothesis more perfectly than seems reasonable. Imagine, for instance, that Prout had advanced his hypothesis and then proceeded to report numerous atomic weights that he had himself measured, each an exact whole number. Such a result looks almost as if it was designed to impress, and just for this reason it fails to.

We may analyse this response as follows: chemists in the early nineteenth century recognized that the measuring techniques available to them were not absolutely precise in their accuracy but were subject to experimental error, and so liable to produce a certain spread of results about the true value. On this assumption, which we label a', it is extremely unlikely that numerous independent atomic weight measurements would all produce exactly whole numbers, even if Prout's hypothesis were true. So $P(e \mid t \,\&\, a')$ is extremely small, and clearly $P(e \mid {\sim}t \,\&\, a')$ could be no larger. Now there are many possible explanations of e, apart from those involving a', one being that the experiments were consciously or unconsciously rigged so as to appear favourable to Prout's hypothesis. If this were the only plausible alternative (and so, in effect, equivalent to ${\sim}a$), $P(e \mid t \,\&\, {\sim}a')$ would be very high, as too $P(e \mid {\sim}t \,\&\, {\sim}a)$. It follows from the equations in section **e,** above that

$$P(e \mid t) \approx P(e \mid t \,\&\, {\sim}a')P({\sim}a') \text{ and}$$

$$P(e \mid {\sim}t) \approx P(e \mid {\sim}t \,\&\, {\sim}a')P({\sim}a')$$

and hence,

$$P(e) \approx P(e \mid t \,\&\, {\sim}a')P({\sim}a')P(t) + P(e \mid {\sim}t \,\&\, {\sim}a')P({\sim}a')P({\sim}t).$$

Now presumably the rigging of the results to produce exactly whole numbers would be equally effective whether t was true or not; in other words,

$$P(e \mid t \,\&\, {\sim}a') = P(e \mid {\sim}t \,\&\, {\sim}a').$$

Therefore,

$$P(t \mid e) = \frac{P(e \mid t)P(t)}{P(e)} \approx \frac{P(e \mid t \,\&\, {\sim}a')P({\sim}a')P(t)}{P(e \mid t \,\&\, {\sim}a')P({\sim}a')} = P(t).$$

Thus e does not confirm t significantly, even though, in a misleading sense, it fits the theory perfectly. This is why it is said to be too good to be true. A similar calculation shows that the probability of a' is diminished, and on the assumptions we have made, this implies that the idea that the experiments were fabricated is rendered more probable. (The above analysis is essentially due to Dorling 1996.)

A famous case of data that were alleged to be too good to be true is that of Mendel's plant-breeding results. Mendel's genetic theory of inheritance allows one to calculate the probabilities of different plants producing specific kinds of offspring. For example, under certain circumstances, pea plants of a certain strain may be calculated to yield round and wrinkled seeds with probabilities 0.75 and 0.25, respectively. Mendel obtained seed frequencies that matched the corresponding probabilities in this and in similar cases remarkably well, suggesting (misleadingly, Fisher contended) substantial support for the genetic theory. Fisher did not believe that Mendel had deliberately falsified his results to appear in better accord with his theory than they really were. To do so, Fisher said, "would contravene the weight of the evidence supplied in detail by . . . [Mendel's] paper as a whole". But Fisher thought it a "possibility among others that Mendel was deceived by some assistant who knew too well what was expected" (1936, p. 132), an explanation that he backed up with some, rather meagre, evidence. Dobzhansky (1967, p. 1589), on the other hand, thought it "at least as plausible" that Mendel had himself discarded results that deviated much from his ideal, in the sincere belief that they were contaminated or that some other accident had befallen them. (For a comprehensive review see Edwards 1986.)

The argument put forward earlier to show that too-exactly whole-number atomic weight measurements would not have supported Prout's hypothesis depends on the existence of some sufficiently plausible alternative hypothesis that would explain the data better. We believe that in general, data are too good to be true relative to one hypothesis only if there are such alternatives. This principle implies that if the method of eliciting atomic weights had long been established as precise and accurate, and if careful precautions had been taken against experimenter bias and deception, so that all the natural alternatives to Prout's hypothesis could be discounted, the inductive force of the data would then no longer be suspicious. Fisher, however, did not subscribe to the principle, at least, not explicitly; he believed that Mendel's results told against the genetic theory, irrespective of any alternative explanations that might be suggested. But despite this official position, Fisher did in fact, as we have just indicated, sometimes appeal to such alternatives when he formulated his argument. We refer again to Fisher's case against Mendel in the next chapter, section **b**.

4.g Ad Hoc Hypotheses

We have been discussing the circumstances in which an important scientific hypothesis, in combination with others, makes a false prediction and yet emerges with its reputation more or less intact, while one or more of the auxiliary hypotheses are largely discredited. We argued that this process necessarily calls for alternatives to the discredited hypotheses to be contemplated. Philosophers, such as Popper and Lakatos, who deny any inductive role for evidence, and who oppose, in particular, the Bayesian approach take note of the fact that scientists often do deal with particular instances of the Duhem problem by proposing alternative hypotheses; some of these philosophers have suggested certain normative rules that purport to say when such alternatives are acceptable and when they are not. Their idea is that a theory that was introduced ad hoc, that is, "for the sole purpose of saving a hypothesis seriously threatened by adverse evidence" (Hempel 1966, p. 29), is in some way inferior. The adhocness idea was largely inspired by certain types of scientific example, which appeared to endorse it, but in our view, the examples are misinterpreted and the idea badly flawed. The following are four such examples.

1 Velikovsky, in a daring book called *Worlds in Collision* that attracted a great deal of interest and controversy some years ago, advanced the theory that the Earth has been subject at various stages in its history to cosmic disasters, through near collisions with massive comets. He claimed that one such comet passed close by our planet during the Israelites' captivity in Egypt, causing many of the remarkable events related in the Bible, such as the ten plagues and the parting of the Red Sea, before settling down as the planet Venus. Because the putative cosmic encounter rocked the entire Earth, Velikovsky expected other peoples to have recorded its consequences too, if they kept records at all. But as a matter of fact, many communities around the world failed to note anything out of the ordinary at the time, an anomaly that Velikovsky attributed to a "collective amnesia". He argued that the cataclysms were so terrifying that whole peoples behaved "as if [they had] obliterated impressions that should be unforgettable". There was a need Velikovsky said, to "uncover the vestiges" of these events, "a task not unlike that of overcoming amnesia in a single person" (1950, p. 288).

Individual amnesia is the issue in the next example.

2 Dianetics is a theory that purports to analyse the causes of insanity and mental stress, which it sees as caused by the 'misfiling' of information in unsuitable locations in the brain. By re-filing these 'engrams', it claims, sanity may be restored, composure enhanced and, incidentally, the memory vastly improved. The therapy is long and expensive and few people have been through it and borne out the theory's claims. However, L. Ron Hubbard, the inventor of Dianetics, trumpeted one purported success, and exhibited this person to a large audience, saying that she had a "full and perfect recall of every moment of her life". But questions from the floor ("What did you have for breakfast on October 3rd, 1942?", "What colour is Mr Hubbard's tie?", and the like) soon demonstrated that the hapless woman had a most imperfect memory. Hubbard explained to the dwindling assembly that when she first appeared on the stage and was asked to come forward "now", the word had frozen her in "present time" and paralysed her ability to recall the past. (See Miller 1987.)

3 Investigations into the IQs of different groups of people show that the average levels of measured intelligence vary. Some environmentalists, so-called, attribute low scores primarily to poor social and educational conditions, an explanation that ran into trouble when a large group of Inuit, leading an aimless, poor and drunken existence, were found to score very highly on IQ tests. The distinguished biologist Peter Medawar (1974), in an effort to deflect the difficulty away from the environmentalist thesis, tried to explain this unexpected observation by saying that an "upbringing in an igloo gives just the right degree of cosiness, security and mutual contact to conduce to a good performance in intelligence tests."

In each of these examples, the theory that was proposed in place of the refuted one seems highly unsatisfactory. It is not likely that any of them would have been advanced, save in response to particular anomalies and in order to evade the consequent difficulty, hence the label 'ad hoc'. But philosophers who attach inductive significance to adhocness recognize that the mere fact that the theory was proposed under such circumstances is not by itself grounds for condemnation. For there are examples, like the following, where a theory that was proposed for the sole purpose of dealing with an anomaly was nevertheless very successful.

4 William Herschel, in 1781, discovered the planet Uranus. Astronomers quickly sought to describe the orbit of the new planet in Newtonian terms, taking account of the perturbing influence of the other known planets, and were able to deduce predictions concerning its future positions. But discrepancies between predicted and observed positions of Uranus substantially exceeded the accepted limits of experimental error, and grew year by year. A few astronomers mooted the possibility that the fault lay with Newton's laws but the prevailing opinion was that there must be some unknown planet acting as an extra source of gravitational attraction on Uranus, which ought to be included in the Newtonian calculations. Two astronomers in particular, Adams and Le Verrier, working independently, were convinced of this and using all the known sightings of Uranus, they calculated in a mathematical *tour de force* where the hypothetical planet must be. The hypothesis was ad hoc, yet it was vindicated when careful telescopic observations as well as studies of old astronomical charts

revealed in 1846 the presence of a planet with the anticipated characteristics. The planet was later called Neptune. Newton's theory was saved, for the time being. (See Smart 1947.)

The Adhocness Criteria

Examples like the first three above have suggested to some philosophers that when a theory t, and an auxiliary hypothesis a, are jointly refuted by some evidence, e', then any replacement, of the form t & a', must not only imply e', but should also have some new, 'independent' empirical implications. And examples similar to the fourth have suggested that if the new theory satisfies this condition, then it is a particular virtue if some of the new, independent implications are verified.

These two criteria were anticipated some four hundred years ago, by the great philosopher Francis Bacon, who objected to any hypothesis that is "only fitted to and made to the measure of those particulars from which it is derived". He argued that a hypothesis should be "larger or wider" than the observations that gave rise to it and said that "we must look to see whether it confirms its largeness and wideness by indicating new particulars" (1620, I, 106). Popper (1963, p. 241) advanced the same criteria, laying it down that a "new theory should be *independently testable*. That is to say, apart from explaining all the *explicanda* which the new theory was designed to explain, it must have new and testable consequences (preferably consequences of a *new kind*)." And secondly, he said, the new theory "should pass the independent tests in question". Bacon called hypotheses that did not meet the criteria "frivolous distinctions", while Popper termed them "ad hoc".[10]

---

[10] The first recorded use of the term 'ad hoc' in this context in English was in 1936, in a review of a psychology book, where the reviewer criticized some explanations proffered by the book's author for certain aspects of childish behaviour:

> There s a suspicion of 'ad-hoc-ness' about the 'explanations'. The whole point is that such an account cannot be satisfactory until we can predict the child's movements from a knowledge of the tensions, vectors and valences which are operative, independent of our knowledge of how the child actually behaved. So far we seem reduced to inventing valences, vectors and tensions from a knowledge of the child's behaviour. (Sprott, p. 249; our italics)

Lakatos (1970, p. 175) refined this terminology, calling a theory that failed the first requirement ad hoc_1, and one that failed the second ad hoc_2, intending these, of course, as terms of disapproval. By these criteria, the theories that Velikovsky, Medawar, and Hubbard advanced in response to anomalous data are probably ad hoc_1, for they seem to make no independent predictions, though of course a closer study of those theories might reverse that assessment. The Adams-Le Verrier hypothesis, on the other hand, is ad hoc in neither sense, because it did make new predictions, some of which were verified by telescopic sightings of Neptune. Again, philosophical and intuitive judgment coincides. Nevertheless, the adhocness criteria are unsound.

This unsoundness is evident both on apriori grounds and through counter-examples, some of which we consider now. For instance, suppose one were examining the hypothesis that a particular urn contains only white counters, and imagine an experiment in which a counter is withdrawn from the urn at random and then, after its colour has been noted, replaced; and suppose that in 10,000 repetitions of this operation 4,950, say, of the selected counters were red and the rest white. This evidence clearly refutes the initial hypothesis taken together with the various necessary auxiliary hypotheses, and it is then natural to conclude that, contrary to the original assumption, the urn contains both red and white counters in approximately equal numbers. This inference seems perfectly reasonable, and the revised hypothesis appears well justified by the evidence, *yet there is no independent evidence for it*. And if we let the urn vaporize immediately after the last counter has been inspected, no such independent evidence would be possible. So the hypothesis about the (late) urn's contents is ad $hoc_{1\,\&\,2}$; but for all that, it seems plausible and satisfactory (Howson 1984; Urbach 1991).

Speculating on the contents of an urn is but a humble form of enquiry, but there are many instances in the higher sciences which have the same import. Take the following one from the science of genetics: suppose it was initially proposed or believed that two phenotypic characteristics of a certain plant are inherited in accordance with Mendel's principles, through the agency of a pair of independently acting genes located on different chromosomes. Imagine now that plant-breeding experiments throw up a surprising number of plants carrying both phenotypes, so that the original hypothesis of independence is rejected in favour of the idea that the genes are linked on the same chromosome. Again, the revised theory would be strongly confirmed, and established as acceptable merely on the evidence that discredited its predecessor, without any further, independent evidence. (Fisher 1970, Chapter IX, presented an example of this sort.)

The history of the discovery of Neptune, which we have already discussed, illustrates the same point. Adams estimated the mass of the hypothetical planet and the elements of its orbit by the mathematical technique of least squares applied to all the positional observations available on Uranus. Adams's hypothesis fitted these observations so well that *even before Neptune had been sighted through the telescope or detected on astronomical charts*, its existence was contemplated with the greatest confidence by the leading astronomers of the day. For instance, in his retirement address as president of the British Association, Sir John Herschel, after remarking that the previous year had seen the discovery of a minor planet, went on: "It has done more. It has given us the probable prospect of the discovery of another. We see it as Columbus saw America from the shores of Spain. Its movements have been felt, trembling along the far-reaching line of our analysis, *with a certainty hardly inferior to that of ocular demonstration*". And the Astronomer Royal, Sir George Airy, who was initially inclined to believe that the problem with Uranus would be resolved by introducing a slight adjustment to the Inverse-Square law, spoke of "*the extreme probability* of now discovering a new planet in a very short time" (quoted by Smart, p. 61; our italics). Neptune was indeed discovered within a very short time.

There is a more general objection to the idea that hypotheses are unacceptable if they are ad hoc. Imagine a scientist who is interested in the conjunction of the hypotheses t & a, whose implication e can be checked in an experiment. The experiment is performed with the result e', incompatible with e, and the scientist ventures a new theory t & a', which is consistent with the observations. And suppose that either no new predictions follow or none has been confirmed, so that the new theory is ad hoc.

Imagine that another scientist, working without knowledge of his colleague's labours, also wishes to test t & a, but chooses a different experiment for this purpose, an experiment with only two possible outcomes: either e or $\sim e$. Of course, he obtains the latter, and having done so, must revise the refuted theory, to t & a', say. This scientist now notices that e' follows from the new theory and performs the orthodox experiment to verify the prediction. The new theory can now count a successful prediction to its credit, so it is not ad hoc.

But this is strange. We have arrived at opposite valuations of the very same theory on the basis of the very same observations, breaching at the same time what we previously called the Equivalence Condition and showing that the standard adhocness criteria are inconsistent. Whatever steps might be taken to resolve the inconsistency, it seems to us that one element ought to be removed, namely, the significance that the criteria attach to the order in which the theory and the evidence were thought up by a particular scientist, for this introduces into the principles of theory evaluation considerations concerning the state of scientists' minds that are irrelevant and incongruous in a methodology with pretensions to objectivity. No such considerations enter the corresponding Bayesian evaluations.

The Bayesian approach, incidentally, explains why people often react with instant incredulity, even derision, when certain ad hoc hypotheses are advanced. Is it likely that their amusement comes from perceiving, or even thinking they perceive, that the hypotheses lead to no new predictions? Surely they are simply struck by the utter implausibility of the claims.

Independent Evidence

The adhocness criteria are formulated in terms that refer to 'independent' evidence, yet this notion is always left vague and intuitive. How can it be made more precise? Probabilistic independence cannot fit the case. For suppose theory h was advanced in response to a refutation by e' and that h both explains that evidence and makes the novel prediction e''. It is the general opinion, certainly shared by Popperians, and also a consequence of Bayes's theorem, that e'' confirms h, provided it is sufficiently improbable, relative to already available information. As discussed earlier in this chapter, such confirmation occurs, in particular, when $P(e'' \mid h \,\&\, e') > P(e'' \mid e')$. But this inequality can hold without e'' and e' being independent in the probabilistic sense.

Logical independence is also not the point here, for e'' might be independent from e' in this sense through some trivial difference, say, by relating to a slightly different place or moment of time. And in that case, e'' would not necessarily confirm or add credibility to h. For, as is intuitive, new evidence supports a theory significantly only when it is significantly different from known results, not just trivially different in the logical sense described. It is this intuition that appears to underlie the idea of independence used in the adhocness criteria.

That 'different' or 'varied' evidence supports a hypothesis more than a similar volume of homogeneous evidence is an old and widely held idea. As Hempel (1966, p. 34) put it: "the confirmation of a hypothesis depends not only on the quantity of the favourable evidence available, but also on its variety: the greater the variety, the stronger the resulting support". So, for example, a report that a stone fell to the ground from a certain height in such-and-such time on a Tuesday is similar to that relating to the stone's fall on a Friday; it is very different, however, from evidence of a planet's trajectory or of a fluid's rise in a particular capillary tube. But although it is often easy enough to classify particular bodies of evidence as either similar or varied, it is not easy to give the notions a precise analysis, except, in our view, in probabilistic terms, in the context of Bayesian induction.

The similar instances in the above list are such that when one of them is known, any other would be expected with considerable confidence. This recalls Francis Bacon's characterisation of similarity in the context of inductive evidence. He spoke of observations "with a promiscuous resemblance one to another, insomuch that if you know one you know all" and was probably the first to point out that it is superfluous to cite more than a small, representative sample of such observations in evidence (see Urbach 1987, pp. 160–64). We are not concerned to give an exhaustive analysis of the intuitive notion, which is probably too vague for that to be possible, but are interested in that aspect of evidential similarity that is pertinent to confirmation. Bacon's observations seem to capture this aspect and we may interpret his idea in probabilistic terms by saying that if two items of evidence, e_2 and e_1, are similar, then $P(e_2 \mid e_1) \approx 1$; when this condition holds, e_2 provides little support for any hypothesis if e_1 has already been cited as evidence. When the pieces of evidence are dissimilar, then $P(e_2 \mid e_1)$ is significantly less than 1, so that e_2 now does add a useful amount of confirmation to any already supplied by e_1. Clearly this characterization allows for similarity to be analysed in terms of degree.

To summarize, the non-Bayesian way of appraising hypotheses, and thereby of solving the Duhem problem, through the notion of adhocness is ungrounded in epistemology, has highly counter-intuitive consequences, and relies on a concept of independence amongst items of evidence that seems unsusceptible to analysis, except in Bayesian terms. In brief, it is not a success.

4.h Designing Experiments

Why should anyone go to the trouble and expense of performing a new experiment and of seeking new evidence? The question has been debated recently. For example, Maher (1990) argues that since evidence can neither conclusively verify nor conclusively refute a theory, Popper's scientific aims cannot be served by gathering fresh data. And since a large part of scientific activity is devoted to that end, if Maher is right, this would constitute yet another serious criticism of Popper's philosophy. Of more concern to us is Miller's claim (1991, p. 2) that Bayesian philosophy comes up against the same difficulty:

> If e is the agent's total evidence, then $P(h \mid e)$ is the value of his probability and that is that. What incentive does he have to change it, for example by obtaining more evidence than he has already? He might do so, enabling his total evidence to advance from e to e^{+}; but in no clear way would $P(h \mid e')$ be a better evaluation of probability than $P(h \mid e)$ was.

But the purpose of a scientific investigation, in the Bayesian view, is not to better evaluate inductive probabilities. It is to diminish uncertainty about a certain aspect of the world.

Suppose the question of interest concerns some parameter. You might start out fairly uncertain about its value, in the sense that your probability distribution over its range of possible values is fairly diffuse. A suitable experiment, if successful, would furnish evidence to lessen that uncertainty by changing the probability distribution, via Bayes's theorem, making it now more concentrated in a particular region; the greater the concentration and the smaller the region the better. This criterion has been given a precise expression by Lindley (1956), in terms of Shannon's characterization of information, and is discussed further in Howson 2002. Lindley showed that in the case where knowledge of a parameter θ is sought, provided the density of x varies with θ, any experiment in which x is measured has an expected yield in information. But, of course, this result is compatible with a well-designed experiment (with a high expected information yield) being disappointingly uninformative in a particular case; and by the same token, a poor experiment may be surprisingly productive of information.

In deciding whether to perform a particular experiment, at least three other factors should be taken into account: the cost of the experiment; the morality of performing it; and the value, both theoretical and practical, of the hypotheses one is interested in. Bayes's theorem, of course, cannot help here.

4.i Under-Determination and Prior Probabilities

We pointed out in Chapter 1 that any data are explicable by infinitely many, mutually incompatible theories, a situation that some philosophers have called the 'under-determination' of theories by data. For example, Galileo carried out numerous experiments on freely falling bodies, in which he examined how long they took to descend various distances. His results led him to propound the well-known law: $s = a + ut + \frac{1}{2}gt^2$, where s is the distance fallen by the body in time t, and a, u and g are constants. Jeffreys (1961, p. 3) pointed out that without contradicting his own experimental results, Galileo might instead have advanced as his law:

$$s = a + ut + \tfrac{1}{2}gt^2 + f(t)(t - t_1)(t - t_2) \ldots (t - t_n),$$

where $t_1, t_2, \ldots, t_n$ are the elapsed times of fall that Galileo recorded in each of his experiments; a, u and g have the same values as above; and f is any function that is not infinite at any of the values $t_1, t_2, \ldots, t_n$. Jeffreys's modification therefore represents an infinity of alternatives to the orthodox theory, all implying Galileo's data, all mutually contradictory, and all making different predictions about future experiments.

There is a similar example due to Goodman (1954; for a lively and illuminating discussion, see Jeffrey 1983, pp. 187–190). He noted that the evidence of many green emeralds, under varied circumstances, would suggest to most observers that all emeralds are green; but he pointed out that that hypothesis bears the same relation to the evidence as does a type of hypothesis that he formulated as 'All emeralds are grue'. Goodman defined something as 'grue' when it was either observed before the present time ($T = 0$) and was green, or was not observed before that time and was blue. Clearly there are infinitely many grue-type predicates and infinitely many corresponding hypotheses, each associated with a different value of $T > 0$. All the current evidence of green emeralds is implied by both the green-hypothesis and the grue variants, yet not more than one of the hypotheses could be true.

As Jeffreys put it, there is always "an infinite number of rules that have held in all previous cases and cannot possibly all hold in future ones." This is a problem for those non-Bayesian scientific methods that regard a theory's scientific value as determined just by $P(e \mid h)$ and, in some versions, by $P(e)$. Such philosophical approaches, of which Popper's is one example, and maximum-likelihood estimation (Chapter 7, section **e**) another, would have to regard the standard law of free fall and Jeffreys's peculiar alternatives as equally good scientific theories relative to the evidence that was available to Galileo, and similarly with Goodman's strange hypotheses concerning emeralds, although these are judgments with which no scientist would agree.

In the Bayesian scheme, if two theories explain the evidence equally well, in the sense that $P(e \mid h_1) = P(e \mid h_2)$, this simply means that their posterior probabilities are in the same ratio as their priors. So theories, such as the contrived variants of Galileo's law and the Goodman grue-alternatives, which have the same relation to the evidence as the orthodox theories and yet are received with incredulity, must have much lower prior probabilities. The role of prior probabilities also accounts for the important feature of scientific reasoning that scientists often prefer a theory that explains the data imperfectly, in the sense that $P(e \mid h) < 1$, to an alternative that explains them perfectly. This occurs when the better explanatory power of the alternative is offset by its inferior prior probability (Jeffreys 1961, p. 4).

This Bayesian account is of course only partial, for we can provide no general account of the genesis of prior probabilities. In some situations, the prior may simply be the posterior probability derived from earlier results and an earlier prior. Sometimes, when there are no such results, a prior probability may be created through what we know from other sources. Consider, for instance, a theory that makes some assertion about a succession of events in the development of a human society; it might, for example, say that the elasticity of demand for herring is constant over a particular period, or that the surnames of all future British prime ministers and American presidents will start with the letter *B*. These theories could possibly be true, but are immensely unlikely to be so. And the reason for this is that the events they describe are the causal effects of numerous, independent processes, whose separate outcomes are improbable. The probability that all the processes will turn out to favour one of the theories in question is therefore the product of many small probabilities and so is itself very small indeed (Urbach 1987b). But the question of how the probabilities of the causal factors are estimated remains. This could be answered by reference to other probabilities, in which case the question is just pushed one stage back, or else by some different form of reasoning. For instance, the 'simplicity' of a hypothesis has been thought to have an influence on its initial probability. This and other possibilities are discussed in Chapter 9.

4.j Conclusion

The various, mostly familiar aspects of scientific reasoning that we have examined have all shown themselves to correspond naturally to aspects of Bayesian logic, whereas non-Bayesian accounts fail more or less completely. So far, we have concentrated chiefly on deterministic theories. We shall see in the next and following chapters that the Bayesian approach applies equally well to statistical reasoning.

Bibliography

Akaike, H. 1973. Information Theory and an Extension of the Maximum Likelihood Principle. In *Second International Symposium of Information Theory*, eds. B.N. Petrov and F. Csáki (Budapest: Akadémiai Kiadó), 267–281.

Anscombe, F.J. 1963. Sequential Medical Trials. *Journal of the American Statistical Association*, Volume 58, 365–383.

Anscombe, F.J., and R.J. Aumann. 1963. A Definition of Subjective Probability. *Annals of Mathematical Statistics*, Volume 34, 199–205.

Armitage, P. 1975. *Sequential Medical Trials.* Second edition. Oxford: Blackwell.

Atkinson, A.C. 1985. *Plots, Transformations, and Regression.* Oxford: Clarendon.

———. 1986. Comment: Aspects of Diagnostic Regression Analysis. *Statistical Science*, Volume 1, 397–402.

Babbage, C. 1827. Notice Respecting some Errors Common to many Tables of Logarithms. *Memoirs of the Astronomical* Society, Volume 3, 65–67.

Bacon, F. 1994 [1620]. *Novum Organum*. Translated and edited by P. Urbach and J. Gibson. Chicago: Open Court.

Barnett, V. 1973. *Comparative Statistical Inference.* New York: Wiley.

Bartha, P. 2004. Countable Additivity and the de Finetti Lottery. *British Journal for the Philosophy of Science*, Volume 55, 301–323.

Bayes, T. 1958 [1763]. An Essay towards Solving a Problem in the Doctrine of Chances. *Philosophical Transactions of the Royal Society*, Volume 53, 370–418. Reprinted with a biographical note by G.A. Barnard in *Biometrika* (1958), Volume 45, 293–315.

Belsley, D.A., E. Kuh, and R.E. Welsch. 1980. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity.* New York: Wiley.

Bernoulli, D. 1738. Specimen theoriae novae de mensura sortis. *Commentarii academiae scientiarum imperialis Petropolitanae*, Volume V, 175–192

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

Berry, D.A. 1989. Ethics and ECMO. *Statistical Science,* Volume 4, 306–310.

Blackwell, D., and L. Dubins. 1962. Merging of Opinions with Increasing Information. *Annals of Mathematical Statistics,* Volume 33, 882–87.

Bland, M. 1987. *An Introduction to Medical Statistics.* Oxford: Oxford University Press.

Blasco, A. 2001. The Bayesian Controversy in Animal Breeding. *Journal of Animal Science*, Volume 79, 2023–046.

Bovens, L. and S. Hartmann. 2003. *Bayesian Epistemology.* Oxford: Oxford University Press.

Bourke, G.J., L.E. Daly, and J. McGilvray. 1985. *Interpretation and Uses of Medical Statistics.* 3rd edition. St. Louis: Mosby.

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

Bradley, R. 1998. A Representation Theorem for a Decision Theory with Conditionals. *Synthese*, Volume 116, 187–229.

Brandt, R. 1986. 'Comment' on Chatterjee and Hadi (1986). *Statistical Science,* Volume 1, 405–07.

Broemeling, L.D. 1985. *Bayesian Analysis of Linear Models.* New York: Dekker.

Brook, R.J. and G.C. Arnold. 1985. *Applied Regression Analysis and Experimental Design.* New York: Dekker.

Burnham, K.P. and D.R. Anderson. 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretical Approach.* New York: Springer-Verlag.

Byar, D.P. *et al.* (seven co-authors). 1976. Randomized Clinical Trials. *New England Journal of Medicine*, 74–80.

Byar, D.P. *et al.* (22 co-authors). 1990. Design Considerations for AIDS Trials. *New England Journal of Medicine*, Volume 323, 1343–48.

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

Casscells W., A. Schoenberger, and T. Grayboys. 1978. Interpretation by Physicians of Clinical Laboratory Results. *New England Journal of Medicine*, Volume 299, 999–1000.

Chatterjee, S., and A.S. Hadi. 1986. Influential Observations, High Leverage Points, and Outliers in Linear Regression. *Statistical Science,* Volume 1, 379–416.

Chatterjee, S., and B. Price. 1977. *Regression Analysis by Example.* New York: Wiley.

Chiang, C.L. 2003. *Statistical Methods of Analysis.* World Scientific Publishing.

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

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

Cook, R.D. 1986. Comment on Chatterjee and Hadi 1986. *Statistical Science,* Volume 1, 393–97.

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

Cox, D.R. 1968. Notes on Some Aspects of Regression Analysis. *Journal of the Royal Statistical Society,* Volume 131A, 265–279.

Cox, R.T. 1961. *The Algebra of Probable Inference.* Baltimore: The Johns Hopkins University Press.

Cramér, H. 1946. *Mathematical Methods of Statistics.* Princeton: Princeton University Press.

Daniel, C., and F.S. Wood. 1980. *Fitting Equations to Data.* New York: Wiley.

David, F.N. 1962. *Games, Gods, and Gambling*. London: Griffin.

Dawid, A.P. 1982. The Well-Calibrated Bayesian. *Journal of the American Statistical Association,* Volume 77, 605–613.

Diaconis, P., and S.L. Zabell. 1982. Updating Subjective Probability. *Journal of the American Statistical Association,* Volume 77, 822–830.

Dobzhansky, T. 1967. Looking Back at Mendel's Discovery. *Science,* Volume 156, 1588–89.

Dorling, J. 1979. Bayesian Personalism, the Methodology of Research Programmes, and Duhem's Problem. *Studies in History and Philosophy of Science,* Volume 10, 177–187.

———. 1996. Further Illustrations of the Bayesian Solution of Duhem's Problem. http://www.princeton.edu/~bayesway/Dorling/dorling.html

Downham, J., ed. 1988. *Issues in Political Opinion Polling.* London: The Market Research Society. Occasional Papers on Market Research.

Duhem, P. 1905. *The Aim and Structure of Physical Theory.* Translated by P.P. Wiener, 1954. Princeton: Princeton University Press.

Dunn, J.M., and G. Hellman. 1986. Dualling: A Critique of an Argument of Popper and Miller. *British Journal for the Philosophy of Science,* Volume 37, 220–23.

Earman, J. 1992. *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory.* Cambridge, Massachusetts: MIT Press.

Edwards, A.L. 1984. *An Introduction to Linear Regression and Correlation.* Second edition. New York: Freeman.

Edwards, A.W.F. 1972. *Likelihood.* Cambridge: Cambridge University Press.

———. 1986. Are Mendel's Results Really Too Close? *Biological Reviews of the Cambridge Philosophical Society,* Volume 61, 295–312.

Edwards, W. 1968. Conservatism in Human Information Processing. In *Formal Representation of Human Judgment,* B. Kleinmuntz, ed., 17–52.

Edwards, W., H. Lindman, and L.J. Savage. 1963. Bayesian Statistical Inference for Psychological Research. *Psychological Review,* Volume 70, 193–242.

Ehrenberg, A.S.C. 1975. *Data Reduction: Analysing and Interpreting Statistical Data.* London: Wiley.

FDA. 1988. *Guideline for the Format and Content of the Clinical and Statistical Sections of New Drug Applications.* Rockville: Center for Drug Evaluation and Research, Food and Drug Administration.

Feller, W. 1950. *An Introduction to Probability Theory and its Applications*, Volume 1. Third edition. New York: Wiley.

Feyerabend, P. 1975. *Against Method.* London: New Left Books.

Finetti, B. de. 1937. La prévision; ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré,* Volume 7, 1–68. Reprinted in 1964 in English translation as 'Foresight: Its Logical Laws, its Subjective Sources', in *Studies in Subjective Probability,* edited by H.E. Kyburg, Jr., and H.E. Smokler (New York: Wiley).

———. 1972. *Probability, Induction, and Statistics*, New York: Wiley.

———. 1974. *Theory of Probability.* Volume 1. New York: Wiley.

Fisher, R.A. 1922. On the Mathematical Foundations of Theoretical Statistics. *Philosophical Transactions of the Royal Society of London,* Volume A222, 309–368.

———. 1930. Inverse Probability. *Proceedings of the Cambridge Philosophical Society*, Volume 26, 528–535.

———. 1935. Statistical Tests. *Nature,* Volume 136, 474.

———. 1936. Has Mendel's Work Been Rediscovered? *Annals of Science,* Volume 1, 115–137.

———. 1947 [1926]. *The Design of Experiments.* Fourth edition. Edinburgh: Oliver and Boyd.

———. 1956. *Statistical Methods and Statistical Inference.* Edinburgh: Oliver and Boyd.

———. 1970 [1925]. *Statistical Methods for Research Workers.* Fourteenth edition. Edinburgh: Oliver and Boyd.

Freeman, P.R. 1993. The Role of *P*-values in Analysing Trial Results. *Statistics in Medicine*, Volume 12, 1433–459.

Gabbay, D. 1994. What Is a Logical System? *What Is a Logical System?*, ed. D. Gabbay, Oxford: Oxford University Press, 179-217.

Gaifman, H. 1964. Concerning Measures in First Order Calculi. *Israel Journal of Mathematics*, Volume 2, 1–18.

———. 1979. Subjective Probability, Natural Predicates, and Hempel's Ravens. *Erkenntnis*, Volume 14, 105-159.

Gaifman, H., and M. Snir. Probabilities over Rich languages, Testing and Randomness. *Journal of Symbolic Logic* 47, 495–548.

Giere, R.N. 1984. *Understanding Scientific Reasoning*. Second edition. New York: Holt, Rinehart.

Gigerenzer, G. 1991. How to Make Cognitive Illusions Disappear: Beyond Heuristics and Biases. *European Review of Social Psychology*, Volume 3, 83–115.

Gillies, D.A. 1973. *An Objective Theory of Probability*. London: Methuen.

———. 1989. Non-Bayesian Confirmation Theory and the Principle of Explanatory Surplus. *Philosophy of Science Association 1988*, edited by A. Fine and J. Loplin, Volume 2 (Pittsburgh: Pittsburgh University Press), 373–381.

———. 1990. Bayesianism versus Falsificationism. *Ratio*, Volume 3, 82–98.

———. 2000. *Philosophical Theories of Probability*. London: Routledge.

Girotto, V., and M. Gonzalez. 2001. Solving Probabilistic and Statistical Problems: A Matter of Information Structure and Question Form. *Cognition*, Volume 78, 247-276.

Glymour, C. 1980. *Theory and Evidence*. Princeton: Princeton University Press.

Good, I.J. 1950. *Probability and the Weighing of Evidence*. London: Griffin.

———. 1961. The Paradox of Confirmation. *British Journal for the Philosophy of Science*, Volume 11, 63–64.

———. 1965. *The Estimation of Probabilities*. Cambridge, Massachusetts: MIT Press.

———. 1969. Discussion of Bruno de Finetti's Paper 'Initial Probabilities: A Prerequisite for any Valid Induction'. *Synthese*, Volume 20, 17–24.

———. 1981. Some Logic and History of Hypothesis Testing. In *Philosophical Foundations of Economics*, edited by J.C. Pitt (Dordrecht: Reidel).

———. 1983. Some History of the Hierarchical Bayes Methodology. *Good Thinking*. Minneapolis: University of Minnesota Press, 95–105.

Goodman, N. 1954. *Fact, Fiction, and Forecast.* London: Athlone.

Gore, S.M. 1981. Assessing Clinical Trials: Why Randomize? *British Medical Journal,* Volume 282, 1958–960.

Grünbaum, A. 1976. Is the Method of Bold Conjectures and Attempted Refutations *Justifiably* the Method of Science? *British Journal for the Philosophy of Science,* Volume 27, 105–136.

Gumbel, E.J. 1952. On the Reliability of the Classical Chi-Square Test. *Annals of Mathematical Statistics,* Volume 23, 253–263.

Gunst, R.F., and R.C. Mason. 1980. *Regression Analysis and its Application.* New York: Dekker.

Hacking, I. 1965. *Logic of Statistical Inference.* Cambridge: Cambridge University Press.

———. 1967. Slightly More Realistic Personal Probability. *Philosophy of Science,* Volume 34, 311–325.

———. 1975. *The Emergence of Probability.* Cambridge: Cambridge University Press.

Halmos, P. 1950. *Measure Theory.* New York: Van Nostrand.

Halpern, J.Y. 1999. Cox's Theorem Revisited. *Journal of Artificial Intelligence Research,* Volume 11, 429–435.

Hays, W.L. 1969 [1963]. *Statistics.* London: Holt, Rinehart and Winston.

Hays, W.L., and R.L. Winkler. 1970. *Statistics: Probability, Inference, and Decision,* Volume 1. New York: Holt, Rinehart.

Hellman, G. 1997. Bayes and Beyond. *Philosophy of Science*, Volume 64.

Hempel, C.G. 1945. Studies in the Logic of Confirmation. *Mind,* Volume 54, 1–26, 97–121. Reprinted in Hempel 1965.

———. 1965. *Aspects of Scientific Explanation.* New York: The Free Press.

———. 1966. *Philosophy of Natural Science.* Englewood Cliffs: Prentice-Hall.

Hodges, J.L., Jr., and E.L. Lehmann. 1970. *Basic Concepts of Probability and Statistics.* Second edition. San Francisco: Holden-Day.

Horwich, P. 1982. *Probability and Evidence.* Cambridge: Cambridge University Press.

———. 1984. Bayesianism and Support by Novel Facts. *British Journal for the Philosophy of Science,* Volume 35, 245–251.

Howson, C. 1973. Must the Logical Probability of Laws be Zero? *British Journal for the Philosophy of Science,* Volume 24, 153–163.

Howson, C., ed. 1976. *Method and Appraisal in the Physical Sciences.* Cambridge: Cambridge University Press.

———. 1987. Popper, Prior Probabilities, and Inductive Inference. *British Journal for the Philosophy of Science,* Volume 38, 207–224.

———. 1988a. On the Consistency of Jeffreys's Simplicity Postulate, and its Role in Bayesian Inference. *Philosophical Quarterly,* Volume 38, 68–83.

———. 1988b. Accommodation, Prediction, and Bayesian Confirmation Theory. *PSA 1988*. A. Fine and J. Leplin, eds., 381–392.

———. 1997. *Logic With Trees*. London: Routledge.

———. 2000. *Hume's Problem: Induction and the Justification of Belief.* Oxford: Clarendon.

———. 2002. Bayesianism in Statistics. *Bayes's Theorem*, ed. R. Swinburne, The Royal Academy: Oxford University Press, 39–71.

Hume, D. 1739. *A Treatise of Human Nature,* Books 1 and 2. London: Fontana.

———. 1777. *An Enquiry Concerning Human Understanding*. Edited by L.A. Selby-Bigge. Oxford: Clarendon.

Jaynes, E.T. 1968. Prior Probabilities. *Institute of Electrical and Electronic Engineers Transactions on Systems Science and Cybernetics,* SSC-4, 227–241.

———. 1973. The Well-Posed Problem. *Foundations of Physics*, Volume 3, 413–500.

———. 1983. *Papers on Probability, Statistics, and Statistical Physics,* edited by R. Rosenkrantz. Dordrecht: Reidel.

———. 1985. Some Random Observations. *Synthese,* Volume 63, 115–138.

———. 2003. *Probability Theory: The Logic of Science*. Cambridge: Cambridge University Press.

Jeffrey, R.C. 1970. 1983. *The Logic of Decision.* Second edition. Chicago: University of Chicago Press.

———. 2004. *Subjective Probability: The Real Thing*. Cambridge: Cambridge University Press.

Jeffreys, H. 1961. *Theory of Probability.* Third edition. Oxford: Clarendon.

Jennison, C., and B.W. Turnbull. 1990. Statistical Approaches to Interim Monitoring: A Review and Commentary. *Statistical Science,* Volume 5, 299–317.

Jevons, W.S. 1874. *The Principles of Science.* London: Macmillan.

Joyce, J.M. 1998. A Nonpragmatic Vindication of Probabilism. *Philosophy of Science*, Volume 65, 575–603.

———. 1999. *The Foundations of Causal Decision Theory*. Cambridge: Cambridge University Press.

Kadane, J., *et al.* 1980. Interactive Elicitation of Opinion for a Normal Linear Model. *Journal of the American Statistical Association,* Volume 75, 845–854.

Kadane, J.B., M.J. Schervish, and T. Seidenfeld. 1999. *Rethinking the Foundations of Statistics*. Cambridge: Cambridge University Press.

Kadane, J.B. and T. Seidenfeld. 1990. Randomization in a Bayesian Perspective. *Journal of Statistical Planning and Inference*, Volume 25, 329–345.

Kant, I. 1783. *Prolegomena to any Future Metaphysics*. Edited by L.W. Beck, 1950. Indianapolis: Bobbs-Merrill.

Kempthorne, O. 1966. Some Aspects of Experimental Inference. *Journal of the American Statistical Association,* Volume 61, 11–34.

———. 1971. Probability, Statistics, and the Knowledge Business. In *Foundations of Statistical Inference,* edited by V.P. Godambe and D.A. Sprott. Toronto: Holt, Rinehart and Winston of Canada.

———. 1979. *The Design and Analysis of Experiments.* Huntington: Robert E. Krieger.

Kendall, M.G., and A. Stuart. 1979. *The Advanced Theory of Statistics,* Volume 2. Fourth edition. London: Griffin.

———. 1983. *The Advanced Theory of Statistics,* Volume 3. Fourth edition. London: Griffin.

Keynes, J.M. 1921. *A Treatise on Probability.* London: Macmillan.

Kieseppä, I.A. 1997. Akaike Information Criterion, Curve-fitting, and the Philosophical Problem of Simplicity. *British Journal for the Philosophy of Science*, Volume 48, 21–48.

Kitcher, P. 1985. *Vaulting Ambition.* Cambridge, Massachusetts: MIT Press.

Kolmogorov, A.N. 1950. *Foundations of the Theory of Probability.* Translated from the German of 1933 by N. Morrison. New York: Chelsea Publishing. Page references are to the 1950 edition.

Korb, K.B. 1994. Infinitely Many Resolutions of Hempel's Paradox. In *Theoretical Aspects of Reasoning about Knowledge*, 138–49, edited by R. Fagin. Asilomar: Morgan Kaufmann.

Kuhn, T.S. 1970 [1962]. *The Structure of Scientific Revolutions.* Second edition. Chicago: University of Chicago Press.

Kyburg, H.E., Jr., and E. Smokler, eds. 1980. *Studies in Subjective Probability.* Huntington: Krieger.

Lakatos, I. 1963. Proofs and Refutations. *British Journal for the Philosophy of Science*, Volume 14, 1–25, 120–139, 221–143, 296, 432.

———. 1968. Criticism and the Methodology of Scientific Research Programmes. *Proceedings of the Aristotelian Society,* Volume 69, 149–186.

———. 1970. Falsification and the Methodology of Scientific Research Programmes. In *Criticism and the Growth of Knowledge*, edited by I. Lakatos and A. Musgrave. Cambridge: Cambridge University Press.

———. 1974. Popper on Demarcation and Induction. In *The Philosophy of Karl Popper*, edited by P.A. Schilpp. La Salle: Open Court.

———. 1978. *Philosophical Papers.* Two volumes. Edited by J. Worrall and G. Currie. Cambridge: Cambridge University Press.

Laplace, P.S. de. 1820. *Essai Philosophique sur les Probabilités.* Page references are to *Philosophical Essay on Probabilities,* 1951. New York: Dover.

Lee, P.M. 1997. *Bayesian Statistics*. Second edition. London: Arnold.

Lewis, D. 1981. A Subjectivist's Guide to Objective Chance. In *Studies in Inductive Logic and Probability,* edited by R.C. Jeffrey, 263–293. Berkeley: University of California Press.

Lewis-Beck, M.S. 1980. *Applied Regression.* Beverley Hills: Sage.

Li, M. and P.B.M. Vitanyi. 1997. *An Introduction to Kolmogorov Complexity Theory and its Applications.* Second edition. Berlin: Springer.

Lindgren, B.W. 1976. *Statistical Theory.* Third edition. New York: Macmillan.

Lindley, D.V. 1957. A Statistical Paradox. *Biometrika,* Volume 44, 187–192.

———. 1965. *Introduction to Probability and Statistics, from a Bayesian Viewpoint.* Two volumes. Cambridge: Cambridge University Press.

———. 1970. Bayesian Analysis in Regression Problems. In *Bayesian Statistics,* edited by D.L. Meyer and R.O. Collier. Itasca: F.E. Peacock.

———. 1971. *Bayesian Statistics: A Review.* Philadelphia: Society for Industrial and Applied Mathematics.

———. 1982. The Role of Randomization in Inference. *Philosophy of Science Association,* Volume 2, 431–446.

———. 1985. *Making Decisions.* Second edition. London: Wiley.

Lindley, D.V., and G.M. El-Sayyad. 1968. The Bayesian Estimation of a Linear Functional Relationship. *Journal of the Royal Statistical Society,* Volume 30B, 190–202.

Lindley, D.V., and L.D. Phillips. 1976. Inference for a Bernoulli Process (a Bayesian View). *American Statistician,* Volume 30, 112–19.

Mackie, J.L. 1963. The Paradox of Confirmation. *British Journal for the Philosophy of Science,* Volume 38, 265–277.

McIntyre, I.M.C. 1991. Tribulations for Clinical Trials. *British Medical Journal,* Volume 302, 1099–1100.

Maher, P. 1990. Why Scientists Gather Evidence. *British Journal for the Philosophy of Science,* Volume 41, 103–119.

——— 1990. Acceptance Without Belief. *PSA 1990*, Volume 1, eds. A. Fine, M. Forbes, and L. Wessels, 381–392.

———1997. Depragmatized Dutch Book Arguments. *Philosophy of Science*, Volume 64, 291–305.

Mallet, J.W. 1880. Revision of the Atomic Weight of Aluminium. *Philosophical Transactions,* Volume 171, 1003–035.

———. 1893. The Stas Memorial Lecture. In *Memorial Lectures delivered before the Chemical Society 1893-1900.* Published 1901. London. Gurney and Jackson.

Mann, H.B., and A. Wald. 1942. On the Choice of the Number of Intervals in the Application of the Chi-Square Test. *Annals of Mathematical Statistics,* Volume 13, 306–317.

Mayo, D.G. 1996. *Error and the Growth of Experimental Knowledge.* Chicago: University of Chicago Press.

Medawar, P. 1974. More Unequal than Others. *New Statesman,* Volume 87, 50–51.

Meier, P. 1975. Statistics and Medical Experimentation. *Biometrics,* Volume 31, 511–529.

Miller, D. 1991. On the Maximization of Expected Futility. PPE Lectures, Lecture 8. Department of Economics: University of Vienna.

Miller, R. 1987. *Bare-faced Messiah.* London: Michael Joseph.

Mises, R. von. 1939 [1928]. *Probability, Statistics, and Truth.* First English edition prepared by H. Geiringer. London: Allen and Unwin.

———. 1957. Second English edition, revised, of *Probability, Statistics and Truth.*

———. 1964. *Mathematical Theory of Probability and Statistics.* New York: Academic Press.

Mood, A.M. 1950. *Introduction to the Theory of Statistics.* New York: McGraw-Hill.

Mood, A.M., and F.A. Graybill. 1963. *Introduction to the Theory of Statistics.* New York: McGraw-Hill.

Musgrave, A. 1975. Popper and 'Diminishing Returns from Repeated Tests', *Australasian Journal of Philosophy,* Volume 53, 248–253.

Myrvold, W.C. and W.L. Harper. 2002. Model Selection and Scientific Inference. *Philosophy of Science*, Volume 69, S124–134.

Neyman, J. 1935. On the Two Different Aspects of the Representative Method: the Method of Stratified Sampling and the Method of Purposive Selection. Reprinted in Neyman 1967, 98–141.

———. 1937. Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability. *Philosophical Transactions of the Royal Society,* Volume 236A, 333–380.

———. 1941. Fiducial Argument and the Theory of Confidence Intervals. *Biometrika,* Volume 32, 128–150. Page references are to the reprint in Neyman 1967.

———. 1952. *Lectures and Conferences on Mathematical Statistics and Probability.* Second edition. Washington, D.C.: U.S. Department of Agriculture.

———. 1967. *A Selection of Early Statistical Papers of J. Neyman.* Cambridge: Cambridge University Press.

Neyman, J., and E.S. Pearson. 1928. On the Use and the Interpretation of Certain Test Criteria for Purposes of Statistical Inference. *Biometrika,* Volume 20, 175–240 (Part I), 263–294 (Part II).

———. 1933. On the Problem of the Most Efficient Tests of Statistical Hypotheses. *Philosophical Transactions of the Royal Society,* Volume 231A, 289–337. Page references are to the reprint in Neyman and Pearson's *Joint Statistical Papers* (Cambridge: Cambridge University Press, 1967).

Pais, A. 1982. *Subtle Is the Lord.* Oxford: Clarendon.

Paris, J. 1994. *The Uncertain Reasoner's Companion.* Cambridge: Cambridge University Press.

Paris, J. and A. Vencovskà. 2001. Common Sense and Stochastic Independence. *Foundations of Bayesianism*, eds. D. Corfield and J. Williamson. Dordrecht: Kluwer, 203–241.

Pearson, E.S. 1966. Some Thoughts on Statistical Inference. In *The Selected Papers of E.S. Pearson,* 276–183. Cambridge: Cambridge University Press.

Pearson, K. 1892. *The Grammar of Science.* Page references are to the edition of 1937 (London: Dent).

Peto, R., *et al.* 1988. Randomised Trial of Prophylactic Daily Aspirin in British Male Doctors. *British Medical Journal,* Volume 296, 313–331.

Phillips, L.D. 1973. *Bayesian Statistics for Social Scientists.* London: Nelson.

———. 1983. A Theoretical Perspective on Heuristics and Biases in Probabilistic Thinking. In *Analysing and Aiding Decision,* edited by P.C. Humphreys, O. Svenson, and A. Van. Amsterdam: North Holland.

Pitowsky, I. 1994. George Boole's Conditions of Possible Experience and the Quantum Puzzle. *British Journal for the Philosophy of Science*, Volume 45, 95–127.

Poincaré, H. 1905. *Science and Hypothesis*. Page references are to the edition of 1952 (New York: Dover).

Polanyi, M. 1962. *Personal Knowledge*. Second edition. London: Routledge.

Pollard, W. 1985. *Bayesian Statistics for Evaluation Research: An Introduction*. Beverly Hills: Sage.

Polya, G. 1954. *Mathematics and Plausible Reasoning*, Volumes 1 and 2. Princeton: Princeton University Press.

Popper, K.R. 1959. The Propensity Interpretation of Probability. *British Journal for the Philosophy of Science*, Volume 10, 25–42.

———. 1959a. *The Logic of Scientific Discovery*. London: Hutchinson.

———. 1960. *The Poverty of Historicism*. London: Routledge.

———. 1963. *Conjectures and Refutations*. London: Routledge.

———. 1972. *Objective Knowledge*. Oxford: Oxford University Press.

———. 1983. A Proof of the Impossibility of Inductive Probability. *Nature*, Volume 302, 687–88.

Pratt, J.W. 1962. On the Foundations of Statistical Inference. *Journal of the American Statistical Association*, Volume 57, 269–326.

———. 1965. Bayesian Interpretation of Standard Inference Statements. *Journal of the Royal Statistical Society*, 27B, 169–203.

Pratt, J.W., H. Raiffa, and R. Schlaifer. 1965. *Introduction to Statistical Decision Theory*.

Prout, W. 1815. On the Relation Between the Specific Gravities of Bodies in Their Gaseous State and the Weights of Their Atoms. *Annals of Philosophy*, Volume 6, 321–330. Reprinted in *Alembic Club Reprints*, No. 20, 1932, 25–37 (Edinburgh: Oliver and Boyd).

Prout, W. 1816. Correction of a Mistake in the Essay on the Relations Between the Specific Gravities of Bodies in Their Gaseous State and the Weights of their Atoms. *Annals of Philosophy*, Volume 7, 111–13.

Putnam, H. 1975. *Collected Papers*, Volume 2. Cambridge: Cambridge University Press.

Ramsey, F.P. 1931. Truth and Probability. In Ramsey, *The Foundations of Mathematics and Other Logical Essays* (London: Routledge).

Rao, C.D. 1965. *Linear Statistical Inference and its Applications*. New York: Wiley.

Renyi, A. 1955. On a New Axiomatic Theory of Probability. *Acta Mathematica Academiae Scientiarum Hungaricae*, Volume VI, 285–335.

Rosenkrantz, R.D. 1977. *Inference, Method, and Decision: Towards a Bayesian Philosophy of Science*. Dordrecht: Reidel.

Salmon, W.C. 1981. Rational Prediction. *British Journal for the Philosophy of Science*, Volume 32, 115–125.

Savage, L.J. 1954. *The Foundations of Statistics.* New York: Wiley.

———. 1962. Subjective Probability and Statistical Practice. In *The Foundations of Statistical Inference,* edited by G.A. Barnard and D.R. Cox (New York: Wiley), 9–35.

———. 1962a. A Prepared Contribution to the Discussion of Savage 1962, 88–89, in the same volume.

Schervish, M., T. Seidenfeld, and J.B. Kadane. 1990. State-Dependent Utilities. *Journal of the American Statistical Association*, Volume 85, 840–847.

Schroeder, L.D., D.L. Sjoquist, and P.E. Stephan. 1986. *Understanding Regression Analysis.* Beverly Hills: Sage.

Schwarz, G. 1978. Estimating the Dimension of a Model. *Annals of Statistics*, Volume 6, 461–464.

Schwartz, D., R. Flamant and J. Lellouch. 1980. *Clinical Trials [L'essay therapeutique chez l'homme].* New York: Academic Press. Translated by M.J.R. Healy.

Scott, D. and P. Krauss. 1966. Assigning Probabilities to Logical Formulas. *Aspects of Inductive Logic*, eds. J. Hintikka and P. Suppes. Amsterdam: North Holland, 219–264.

Seal, H.L. 1967. The Historical Development of the Gauss Linear Model. *Biometrika,* Volume 57, 1–24.

Seber, G.A.F. 1977. *Linear Regression Analysis.* New York: Wiley.

Seidenfeld, T. 1979. *Philosophical Problems of Statistical Inference.* Dordrecht: Reidel.

———. 1979. Why I Am Not an Objective Bayesian: Some Reflections Prompted by Rosenkrantz. *Theory and Decision*, Volume 11, 413–440.

Shimony, A. 1970. Scientific Inference. In *Pittsburgh Studies in the Philosophy of Science,* Volume 4, edited by R.G. Colodny. Pittsburgh: Pittsburgh University Press.

———. 1985. The Status of the Principle of Maximum Entropy. *Synthese,* Volume 68, 35–53.

———. 1993 [1988]. An Adamite Derivation of the Principles of the Calculus of Probability. In Shimony, *The Search for a Naturalistic World View*, Volume 1 (Cambridge: Cambridge University Press), 151–162.

Shore, J.E. and R.W. Johnson. 1980. Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy. *IEEE Transactions on Information Theory* 26:1, 26–37.

Skyrms, B. 1977. *Choice and Chance.* Belmont: Wadsworth.

Smart, W.M. 1947. John Couch Adams and the Discovery of Neptune. *Occasional Notes of the Royal Astronomical Society,* No. 11.

Smith, T.M. 1983. On the Validity of Inferences from Non-random Samples. *Journal of the Royal Statistical Society*, Volume 146A, 394–403.

Smullyan, R. 1968. *First Order Logic*. Berlin: Springer.

Sober, E. and M. Forster. 1994. How to Tell When Simpler, More Unified, Or Less Ad Hoc Theories Will Provide More Accurate Predictions. *British Journal for the Philosophy of Science*, Volume 45, 1–37.

Spielman, S. 1976. Exchangeability and the Certainty of Objective Randomness. *Journal of Philosophical Logic*, Volume 5, 399–406.

Sprent, P.1969. *Models in Regression*. London: Methuen.

Sprott, W.J.H. 1936. Review of K. Lewin's *A Dynamical Theory of Personality*. *Mind*, Volume 45, 246–251.

Stachel, J. 1998. *Einstein's Miraculous Year: Five Papers that Changed the Face of Physics*. Princeton: Princeton University Press.

Stas, J.S. 1860. Researches on the Mutual Relations of Atomic Weights. *Bulletin de l'Académie Royale de Belgique*, 208–336. Reprinted in part in *Alembic Club Reprints*, No. 20, 1932 (Edinburgh: Oliver and Boyd), 41–47.

Stuart, A 1954. Too Good to Be True. *Applied Statistics*, Volume 3, 29–32.

———. 1962. *Basic Ideas of Scientific Sampling*. London: Griffin.

Sudbery, A. 1986. *Quantum Mechanics and the Particles of Nature*. Cambridge: Cambridge University Press.

Suzuki, S. 2005. The Old Evidence Problem and AGM Theory. *Annals of the Japan Association for Philosophy of Science*, 1–20.

Swinburne, R.G. 1971. The Paradoxes of Confirmation: A Survey. *American Philosophical Quarterly*, Volume 8, 318–329.

Tanur, J.M., *et al.* 1989. *Statistics: A Guide to the Unknown*. Third Edition. Duxbury Press.

Teller, P. 1973. Conditionalisation and Observation. *Synthese*, Volume 26, 218–258.

Thomson, T. 1818. Some Additional Observations on the Weights of the Atoms of Chemical Bodies. *Annals of Philosophy*, Volume 12, 338–350.

Uffink, J. 1995. Can the Maximum Entropy Method be Explained as a Consistency Requirement? *Studies in the History and Philosophy of Modern Physics*, Volume 26B, 223–261.

Urbach, P. 1981. On the Utility of Repeating the 'Same Experiment'. *Australasian Journal of Philosophy*, Volume 59, 151–162.

———. 1985. Randomization and the Design of Experiments. *Philosophy of Science*, Volume 52, 256–273.

———. 1987. *Francis Bacon's Philosophy of Science.* La Salle: Open Court.

———. 1987a. Clinical Trial and Random Error. *New Scientist,* Volume 116, 52–55.

———. 1987b. The Scientific Standing of Evolutionary Theories of Society. *The LSE Quarterly,* Volume 1, 23–42.

———. 1989. Random Sampling and the Principles of Estimation. *Proceedings of the Aristotelian Society,* Volume 89, 143–164.

———. 1991. Bayesian Methodology: Some Criticisms Answered. *Ratio (New Series),* Volume 4, 170–184.

———. 1992. Regression Analysis: Classical and Bayesian. *British Journal for the Philosophy of Science,* Volume 43, 311–342.

———. 1993. The Value of Randomization and Control in Clinical Trials. *Statistics in Medicine,* Volume 12, 1421–431.

Van Fraassen, B.C. 1980. *The Scientific Image.* Oxford: Clarendon.

———. 1983. Calibration: A Frequency Justification for Personal Probability. In R.S. Cohen and L. Laudan, eds., *Physics, Philosophy, and Psychoanalysis* (Dordrecht: Reidel), 295–321.

———. 1984. Belief and the Will. *Journal of Philosophy,* Volume LXXXI, 235–256.

———. 1989. *Laws and Symmetry.* Oxford: Clarendon.

Velikovsky, I. 1950. *Worlds in Collision.* London: Gollancz. Page references are to the 1972 edition, published by Sphere.

Velleman, P.F. 1986. Comment on Chatterjee, S., and Hadi, A.S. 1986. *Statistical Science,* Volume 1, 412–15.

Velleman, P.F., and R.E. Welsch. 1981. Efficient Computing of Regression Diagnostics. *American Statistician,* Volume 35, 234–242.

Venn, J. 1866. *The Logic of Chance.* London: Macmillan.

Vranas, P.B.M. 2004. Hempel's Raven Paradox: A Lacuna in the Standard Bayesian Solution. *British Journal for the Philosophy of Science*, Volume 55, 545–560.

Wall, P. 1999. *Pain: The Science of Suffering.* London: Weidenfeld and Nicolson.

Watkins, J.W.N. 1985. *Science and Scepticism.* London: Hutchinson and Princeton: Princeton University Press.

———. 1987. A New View of Scientific Rationality. In *Rational Change in Science,* edited by J. Pitt and M. Pera. Dordrecht: Reidel.

Weinberg, S. and K. Goldberg. 1990. *Statistics for the Behavioral Sciences.* Cambridge: Cambridge University Press.

Weisberg, S. 1980. *Applied Linear Regression.* New York: Wiley.

Welsch, R.E. 1986. Comment on Chatterjee, S., and Hadi, A.S. 1986. *Statistical Science,* Volume 1, 403–05.

Whitehead, J. 1993. The Case for Frequentism in Clinical Trials. *Statistics in Medicine,* Volume 12, 1405–413.

Williams, P.M. 1980. Bayesian Conditionalisation and the Principle of Minimum Information. *British Journal for the Philosophy of Science,* Volume 31, 131–144.

Williamson, J. 1999. Countable Additivity and Subjective Probability. *British Journal for the Philosophy of Science*, Volume 50, 401–416.

Williamson, J. 2005. *Bayesian Nets and Causality: Philosophical and Computational Foundations*. Oxford: Oxford University Press.

Williamson, J. and D. Corfield. 2001. Introduction: Bayesianism into the Twenty-First Century. In Corfield, D. and Williamson, J., eds., *Foundations of Bayesianism* (Dordrecht: Kluwer).

Wonnacott, T.H., and R.J. Wonnacott. 1980. *Regression: A Second Course in Statistics.* New York: Wiley.

Wood, M. 2003. *Making Sense of Statistics*. New York: Palgrave Macmillan.

Yates, F. 1981. *Sampling Methods for Censuses and Surveys.* Fourth edition. London: Griffin.