# Interpreting Probability

## Controversies and Developments in the Early Twentieth Century

DAVID HOWIE

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

© David Howie 2002

This book is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 2002

Printed in the United Kingdom at the University Press, Cambridge

*Typeface* Times Roman 10.25/13 pt. *System* $\LaTeX\ 2_{\varepsilon}$ [TB]

*A catalog record for this book is available from the British Library.*

*Library of Congress Cataloging in Publication Data*

Howie, David, 1970–

Interpreting probability : controversies and developments in the early
twentieth century / David Howie.

p. cm. – (Cambridge studies in probability, induction, and decision theory)

Includes bibliographical references and index.

ISBN 0-521-81251-8

1. Probabilities. 2. Bayesian statistical decision theory. 3. Jeffreys, Harold, Sir, 1891–1989
4. Fisher, Ronald Aylmer, Sir, 1890–1962. I. Title. II. Series.

QA273.A4 H69 2002 2001052430

ISBN 0 521 81251 8 hardback

# Contents

# 1

# Introduction

> "It is curious how often the most acute and powerful intellects have gone astray in the calculation of probabilities."
>
> William Stanley Jevons

## 1.1. THE MEANING OF PROBABILITY

The single term 'probability' can be used in several distinct senses. These fall into two main groups. A probability can be a limiting ratio in a sequence of repeatable events. Thus the statement that a coin has a 50% probability of landing heads is usually taken to mean that approximately half of a series of tosses will be heads, the ratio becoming ever more exact as the series is extended. But a probability can also stand for something less tangible: a degree of knowledge or belief. In this case, the probability can apply not just to sequences, but also to single events. The weather forecaster who predicts rain tomorrow with a probability of $\frac{1}{2}$ is not referring to a sequence of future days. He is concerned more to make a reliable forecast for tomorrow than to speculate further ahead; besides, the forecast is based on particular atmospheric conditions that will never be repeated. Instead, the forecaster is expressing his confidence of rain, based on all the available information, as a value on a scale on which 0 and 1 represent absolute certainty of no rain and rain respectively.

The former is called the frequency interpretation of probability, the latter the epistemic or 'degree of belief' or Bayesian interpretation, after the Reverend Thomas Bayes, an eighteenth-century writer on probability.[1] A frequency probability is a property of the world. It applies to chance events. A Bayesian probability, in contrast, is a mental construct that represents uncertainty. It applies not directly to events, but to our knowledge of them, and can

[1] I shall use the terms "Bayesian" and "Bayesianism" when discussing the epistemic interpretation, though they were not common until after World War II.

thus be used in determinate situations. A Bayesian can speak of the probability of a tossed coin, for example, even if he believes that with precise knowledge of the physical conditions of the toss – the coin's initial position and mass distribution, the force and direction of the flick, the state of the surrounding air – he could predict exactly how it would land.

The difference between the two interpretations is not merely semantic, but reflects different attitudes to what constitutes relevant knowledge. A frequentist would take the probability of throwing a head with a biased coin as some value $P$, where $P \neq \frac{1}{2}$. This would express his expectation that more heads than tails, or vice versa, would be found in a long sequence of results. A Bayesian, however, would continue to regard the probability of a head as $\frac{1}{2}$, since unless he suspected the direction of the bias, he would have no more reason to expect the next throw to be a head than a tail. The frequentist would estimate a value for $P$ from a number of trial tosses; the Bayesian would revise his initial probability assessment with each successive result. As the number of trial tosses is increased, the values of the two probabilities will tend to converge. But the interpretations of these values remain distinct.

## 1.2. THE HISTORY OF PROBABILITY

The mathematical theory of probability can be traced back to the mid-seventeenth century. Early probabilists such as Fermat and Pascal focused on games of chance. Here the frequency interpretation is natural: the roll of a die, or the selection of cards from a deck, is a well-defined and repeatable event for which exact sampling ratios can easily be calculated. Yet these men were not just canny gamblers. They intended their theory of probability to account for hazard and uncertainty in everyday life too. This required an inversion of the usual methods. Games of chance involve direct probabilities: from the constitution of a deck of cards, the odds of being dealt a specified hand can be calculated precisely using combinatorial rules. Needed for general situations, however, was the reverse process – working backward from the observation of an event or sequence of events to their generating probabilities, and hence the likely distribution of events in the future.

This problem of 'inverse probability' was addressed by Bayes in a paper published in 1764. His method could be applied to the common model in which draws are made from an urn filled with unknown quantities of differently colored balls. It could be used to calculate the likely proportion of colors in the urn, and hence the probability that the next selected ball will be of a given color, from the information of previous sampling. If the urn is known to contain only black and white balls, for example, and assuming that all possible fractions of black balls in the urn are initially equally likely, and that balls drawn are replaced, then the observation of $p$ black and $q$ white balls from the first $(p + q)$ draws yields the probability that the next ball drawn will be black of $(p + 1)/(p + q + 2)$. The urn models the process of learning from observations. The calculus of inverse probability thus enabled unknown causes to be inferred from known effects.

The method was championed from the late eighteenth century by Laplace as a universal model of rationality. He applied this ‘doctrine of chances’ widely. In the courtroom, for example, the probability that the accused was guilty could be calculated from the known jury majority, and hence used to assess voting procedures. Probabilists such as Laplace, however, tended not to notice that inverting the equations changed the meaning of probability. The probability of picking a ball of a particular color was a function of the relative proportions in the urn. The probability of guilt was different: it was a quantified judgment in the light of a specified body of knowledge.

Inverse probability began to fall from favor around 1840. Its opponents scoffed at the idea that degrees of rational belief could be measured, and pointed out that since the method could be used only to update beliefs rather than derive them *a priori*, it usually relied on probability evaluations that were arbitrary, or at least unjustified. The mid-nineteenth century also saw a rapid growth in attempts to quantify the human subject, as marked by an eruption of published numerical data touching on all matters of life, particularly crime rates and incidence of disease. The stable ratios observed in these new statistics seemed a firmer base for probability than the vague and imperfect knowledge of the inverse method. Nevertheless, Bayesian methods were often the only way to tackle a problem, and many philosophers and scientists continued through Victorian times to maintain that inverse probability, if only in qualitative form, modeled the process by which hypotheses were inferred from experimental or observational evidence.

As a model of the scientific method, inverse probability reached its nadir around 1930. Yet this eclipse was temporary. Following pioneering work by L.J. Savage, Bruno de Finetti, Dennis Lindley, and I.J. Good in the 1950s, Bayesianism is once again in vogue in statistical, economic, and especially philosophical circles. From the late 1960s, experimental psychologists have typically modeled the human mind as Bayesian, and legal scholars have seriously proposed that jurors be schooled in Bayesian techniques. Bayesian confirmation theory is dominant in contemporary philosophy, and even though most scientists continue to rely on the frequency interpretation for data analysis, the Bayesian calculus was recently used in the observation of the top quark.[2]

## 1.3. SCOPE OF THIS BOOK

This is a study of the two types of probability, and an investigation of how, despite being adopted, at least implicitly, by many scientists and statisticians in the eighteenth and nineteenth centuries, Bayesianism was discredited as a theory of scientific inference during the 1920s and 1930s. I shall focus on two British scientists, Sir Harold Jeffreys (1891–1989) and Sir Ronald Aylmer Fisher (1890–1962). Fisher was a biological and agricultural statistician who specialized in problems of inheritance. In addition to developing new statistical methods, he is credited, with J.B.S. Haldane and Sewall Wright, with synthesizing the biometric and Mendelian approaches into a theory of evolution based on gene frequencies in populations. Jeffreys was a theoretical geophysicist and astrophysicist who also made seminal contributions to the fledgling fields of atmospheric physics and seismology. He worked largely alone to construct sophisticated mathematical models of the Earth and the solar system from the principles of classical mechanics. When tested against the available evidence, such models could be used to investigate the origins and internal structure of the Earth; in this way, Jeffreys inferred in 1926 that the Earth's core was liquid. Jeffreys knew that such conclusions were always uncertain, and that new results could force their revision or abandonment, and tried to account for this with a formal theory of scientific reasoning based on Bayesian probability. Fisher, in contrast, was more concerned with the reduction and evaluation of experimental data than an assessment of how likely a theory was to be true. He regarded Bayesian methods as unfounded in principle and misleading in practice, and worked to replace them with a theory of statistical inference based on frequencies.

The two men developed their theories independently during the 1920s, but clashed publicly in the early 1930s over the theory of errors. This was not simply a mathematical matter. Though eighteenth-century probabilists had been careful to attune the doctrine of chances to the intuitions of men of quality, by the beginning of the twentieth century the probability calculus was taken to define the rational attitude in matters of uncertainty. For both Fisher and Jeffreys, therefore, a theory of probability carried a moral imperative. Not only applicable to analysis, it constituted 'correct' scientific judgment and

[2] Bhat, Prosper, and Snyder 1997.

fixed the proper role of the scientist. Fisher believed that a scientist should not step beyond the available data. Consequently, he regarded the Bayesian method as inherently unscientific, because initial probability assignments were arbitrary and hence sanctioned individual prejudice. Jeffreys, in contrast, considered science to be simply a formal version of common sense. Though he was concerned that the calculus reflect the general consensus among scientists, he took it as obvious that the expertise and experience of the individual should be a necessary component of a theory of science.

### 1.4. METHODS AND ARGUMENT

Science has traditionally been presented as a logical and progressive activity in which theoretical models, initially abstracted from facts of observation, are refined or revised with incoming experimental data in order to account for a range of physical phenomena with increasing precision and generality. In the last few decades, however, a number of historians have reacted to this account. The 'scientific method,' they argue, is less a description of usual practice than a fiction used by scientists to bolster confidence in their theories. Science is not an impersonal or faceless activity, but the product of individuals, often with strong political affinities or religious beliefs, and usually eager to win recognition and further their careers.[3] Such concerns can influence research in a number of ways. Though some historical accounts have depicted particular scientific theories as the product of little more than, say, middle-class values or certain ideological commitments,[4] the more persuasive seem to indicate that the effects of such 'external' factors are more subtle. In addition to directing the focus of research, they can shape standards of scientific conduct. Thus various scholars have argued that characteristic approaches to science form around different communities, whether these be marked by distinct experimental apparatus or analytical procedures,[5] or the

[3] An early advocate of this approach was the American sociologist Robert Merton, who from the late 1930s studied the normative structure of science, and how it related both to the wider society and to science as a profession. See the collection of essays, Merton 1983.

[4] MacKenzie 1981 argues that British statisticians' development of new techniques during the nineteenth century was bound up with their middle-class values and commitment to eugenics.

[5] For example, Traweek 1988 relates the experimental practices and career strategies of nuclear physicists to the apparatus of their groups; Buchwald 1994 points to the importance of instrumentation on the interpretation of Hertz's electromagnetic experiments.

wider boundaries associated with disciplines and sub-disciplines,[6] or national groups or even genders.[7] Sometimes, such approaches are borrowed wholesale from more familiar forms of life. Shapin and Schaffer have argued, for example, that the culture of modern science dates to the seventeenth century, and was appropriated from the norms of behavior of the English gentleman.[8] Their study reveals much about how science became authoritative. The credibility of particular results depended on the social status of the individuals who vouched for them. The gentleman could not be doubted without giving him the lie; contrariwise, artisans, or those in employ, could hardly be disinterested. Contrasting with this 'English' style of persuasion is that associated with the nineteenth-century German bureaucratic culture of 'exact sensibility,' which tended to place trust in numerically precise experiments with quantitative error estimations.[9]

Some sociologists have gone further, and argued that in influencing the interpretation of scientific theories, and the criteria by which they are evaluated and tested, such social and cultural factors determine the content of scientific knowledge too.[10] Others have pointed to the holistic character of research, noting that experiments are generally not simple exercises in fact-gathering, but depend on a variety of factors – the equipment, the skills of the experimenter, his theoretical and phenomenological understanding, and so on – that in practice can be reconciled in a number of distinct but consistent ways. Thus an experiment has no single successful outcome, and any particular result must be regarded as local and contingent, in part the choice of the experimentalist, his expectations and presuppositions, and the intellectual and experimental resources available.[11]

[6] For example, Heilbron and Seidel 1989 describe the organizational structure that evolved around particle physics; Bromberg 1991 and Forman 1992 relate the development of the laser and maser, respectively, to the utilitarian needs of the sponsoring government and military institutions.

[7] For example, Pestre and Krige 1992 analyze how national work styles influenced the growth of large-scale physics research; Hunt 1991b describes the influence of the 'British' tradition that the aether must be reducible to mechanical principles on those scientists attempting to explicate Maxwell's electromagnetic theory.

[8] Shapin and Schaffer 1985, Shapin 1994.

[9] See the essays in Wise 1995.

[10] Bloor [1976] 1991 is a key text of the 'sociology of scientific knowledge,' a field much influenced by Kuhn's 1962 theory of scientific revolutions. For a survey of the field, see Barnes, Bloor, and Henry 1996.

[11] Hacking 1992 outlines this holistic theory of scientific research. For examples, see Galison's 1987 study of high-energy physics and Pickering's 1989 study of the weak neutral current.

Such approaches have become notorious in recent years, and have been attacked by scientists as undermining their claim that scientific knowledge is reliable.[12] To recognize that science is not 'value free,' however, is not to regard its knowledge as false. Languages are wholly artificial yet still help to explain and manipulate the world. Most historians and sociologists of science are simply attempting to stimulate new directions in their fields rather than questioning the authority or cognitive content of science.

This book is concerned not so much with experimental science as with *theories* of experimental science. The process of theorizing, however, shares with experimentation a creative and open-ended character. Not even a mathematical theory is a straightforward product of logic. Bare equations mean nothing; they must be interpreted through models and analogies, and to be used in science they require rules to govern how and to what they apply.[13] As Wittgenstein indicated, mathematics too is a form of life. My work examines the case of probabilistic theories of scientific inference. The concept of probability has a degree of plasticity, and I will argue that its interpretation and application depends in part on the particular context.

For Fisher and Jeffreys, a strong claim can be made about the relative influence of the various contextual factors. Although with diametrically opposed views of the role of probability in science, they were otherwise very similar. Almost exact contemporaries, both came from respectable but straitened middle class backgrounds to be educated in Cambridge around the same time. They both had wide-ranging scientific interests, and both were early influenced by the polymath Karl Pearson. Indeed, during the 1940s and 1950s, they lived within a minute's walk of each other in north Cambridge. The chief difference stemmed from another similarity. Both were primarily practicing scientists rather than professional statisticians or mathematicians, and each developed his theory of probability in response to specific features of his research discipline. It was these, I shall argue, that was the major influence for each man in selecting his interpretation of probability.

Fisher worked on Mendelian genetics, one of the few wholly probabilistic theories in natural science, and unique in involving a set of clearly-defined outcomes, each of exact probability. A chance mechanism was central to

[12] See, e.g., Sokal and Bricmont 1998.

[13] For an example along these lines, see Pickering and Stephanides's 1992 account of 'quaternions' – generalized complex numbers developed by William Hamilton for problems in applied mathematics; also see Warwick 1992, 1993, who has considered Einstein's relativity theory as a broad resource that could be selectively mobilized according to the interests of various groups and individuals.

Mendel's theory, and the outcome of an experiment was to be explained not in terms of uncertain causes or incomplete knowledge, but with reference to the irreducible unpredictability of a definite if initially unknown probability. Fisher considered the results of his breeding and agricultural experiments as random samples from a distribution of fixed probability. This probability had a clear, indeed almost palpable, interpretation as the limiting frequency or ratio of, say, a biological trait appearing in a large ensemble of data.

For Jeffreys, such an interpretation was less natural. His research in geophysics did not fit the Mendelian model. From our knowledge of the Earth, what is the probability that its core is molten? It made no sense to talk of an ensemble of Earths. Instead, the probability represented imperfect knowledge. After all, the core was either molten or not. Geophysics involved much inference from incomplete data, but few of the questions that interested Jeffreys were explicable in terms of repeated sampling. Was it reasonable to believe on the current evidence that the solar system formed by condensation, or that there will be higher rainfall next year than this? Jeffreys decided that all scientific questions were essentially of this uncertain character, and that a theory of scientific inference should therefore be based on probability. Such a probability was not a frequency, however, but clearly a degree of knowledge relative to a given body of data.

Disciplinary concerns entered another way too. Like many of the founders of modern statistical theory, Fisher was a keen proponent of eugenics. This issue was obscured by charged and emotive rhetoric, and Fisher believed that an objective and mathematically rigorous way of dealing with quantitative data could remove individual bias and thus render the lessons of eugenics compelling. Moreover, such a method, if suitably general, could also be used to instill proper scientific standards in new academic disciplines, and prescribe correct experimental procedure. Since he was fashioning statistics as a normative procedure, and the statistician as a generalized scientific expert, with duties not merely of analysis, but in the design of experiments to yield unambiguous answers, Fisher regarded the subjectivity and ambiguity of the Bayesian interpretation with abhorrence. For Jeffreys, on the other hand, matters of public policy were not an issue, and thus strict objectivity not at so great a premium. Fisher's recipes for experimental design were not much help either. One did not experiment on the Earth or the solar system, but relied instead on observations from seismic stations or telescopes. Since this data was often sparse or ambiguous, Fisher's brute statistical recipes were useless. Instead, to tease any genuine physical effects from the scatter and noise, the scientist needed to use his personal expertise and previous experience. Fisher would regard this as a subjective matter, but for Jeffreys it was necessary if any inferences were to be drawn from the available evidence, and was precisely the sort of information that could be encoded in the Bayesian model.

Each interpretation of probability, therefore, seemed suited to a particular sort of inquiry. Even so, perhaps Jeffreys was simply wrong and Fisher right, or vice versa? To meet this objection, I shall examine their dispute in detail. Much of the argument was mathematical, seemingly unpromising territory for the contextual historian. Yet as the ‘strong programme’ of the Edinburgh school has established – arguably their most constructive product – the deep assumptions and commitments of a research program, though usually tacit, are often unveiled during periods of controversy.[14] The exchange between Fisher and Jeffreys was one of the few occasions in which a prominent frequentist was forced to confront an intelligent and thoughtful opponent rather than a straw man, and indicates that their versions of frequentism and Bayesianism were incompatible but coherent. Neither scheme was perfect for scientific inference, but each had certain advantages – according to Fisher and Jeffreys’s different ideas of what counted as genuine and trustworthy science – when compared with the other.

The debate between Fisher and Jeffreys ended inconclusively, each man unyielding and confident in his own methods. Yet in the following decade, inverse probability was almost entirely eclipsed as a valid mode of scientific reasoning. I have argued that the clash between Fisher and Jeffreys was not a consequence of error on one side. Why then was Jeffreys’s work unpersuasive? Contemporary scientists tend to be unimpressed with fancy narratives of their discipline; they believe that social or cultural factors intrude to pervert proper scientific behavior rather than constitute it, and retrospectively gloss most scientific disagreements as resulting from inadequate evidence.[15] Yet this reading is not compelling for the clash between the Bayesian and frequentist schools. The mathematics on each side was not in question, any conceptual difficulties with the Bayesian position had been recognized for decades, and there were no new experimental results to resolve the dispute. Instead, it was a particular alignment of the contextual resources used in the 1930s to interpret and evaluate the Bayesian method that ultimately led to its rejection.

One of these was a broad reconception of the limits of knowledge. The epistemic interpretation of probability flourishes in times of determinism. In a mechanistic world there are no real probabilities; instead the concept reflects incomplete knowledge. Thus in statistical physics probabilities provide an approximate description of the motion of vast numbers of atoms in a volume

[14] See, for example, Collins 1985.
[15] See Sokal and Bricmont 1998; also Laudan 1990.

of gas. But scientists changed their view of causation during the 1920s and 1930s. According to the 'new physics' – quantum mechanics – the world was inherently stochastic; radioactive decay was random rather than due to some hidden cause. Of course, an indeterministic theory of science does not force a Bayesian interpretation – Jeffreys argued that the very existence of the new physics meant that theories previously held as conclusive were defeasible, and thus by extension that all scientific generalizations should only be asserted with epistemic probability – but the rise of quantum mechanics accustomed physicists to regard chance not as a consequence of imperfect human inquiry but as a feature of nature, and to explain experimental results in terms of probabilities that could be equated with observed frequencies.

Bayesianism suffered for more prosaic reasons too. Fisher's numerical and objective methods held for social scientists the promise of much-needed rigor and credibility. Indeed, so eager were they to adopt his very particular model of scientific experimentation that they reconfigured some fields, such as psychology and to some extent clinical medicine, to conform with it. And if social scientists were keen to buy objective methods, statisticians were keen to sell. Their rejection of Bayesianism can be seen as part of a strategy of professionalization: by presenting theirs as a general and objective method of analysis, statisticians were justifying their discipline as distinct from any particular field, and thus a candidate for separate university departments, journals, and so on.

The precise combination and influence of such factors differed between disciplines. Some statisticians regarded Fisher's frequentism as no less conceptually vulnerable than Jeffreys's Bayesianism, and rejected both accounts of inference in favor of the severe frequentism of the decision-theory approach developed during the 1930s by Jerzy Neyman and Egon Pearson. Some scientists also argued that a theory of probabilistic inference was not necessary for science. The astronomer Sir Arthur Eddington, the most flamboyant popularizer of Einstein's Theory of Relativity in Britain, asserted loudly that particular general laws must be true on purely *a priori* considerations, and cited Einstein's theory as a case in point. Probability was to be relegated to data analysis. In Germany, the philosopher-physicist Richard von Mises was developing his own frequentist interpretation of probability, but also regarded the concept as having only a restricted place in scientific inquiry. Most mathematicians, though still loosely frequentist, preferred from 1933 to follow the Russian Andrei Kolmogorov with an axiomatic definition of probability based on measure theory. Philosophers persisted with a broadly epistemic probability, but turned for its exposition not to Jeffreys but to John Maynard Keynes's logical interpretation of 1921, or occasionally to that of the mathematician

Frank Ramsey, who in 1926 had sketched the probability calculus as rules for the consistent assignment of degrees of belief, in practice measured by betting odds.

As the concept of probability fragmented, so Fisher and Jeffreys, having battled from opposite sides of a seemingly sharp boundary, found themselves jostled ever closer. In his later papers, Fisher adopted subjectivist language to defend the pragmatic aspects of his own theory against the Neyman–Pearson school. Jeffreys too saw Fisher's approach as more level headed than the alternative versions of non-scientists; soon after their exchange, Jeffreys wrote that his and Fisher's methods almost invariably led to the same conclusions, and where they differed, both were doubtful. Indeed, the Bayesian revival of the 1950s was based not so much on Jeffreys's work as on a more extreme version – personalism – developed, also during the 1930s, by the Italian Bruno de Finetti. De Finetti rejected Jeffreys's attempts to produce standard prior probabilities and his definition of a probability as a degree of knowledge or rational belief. A probability no longer related a hypothesis to a body of data, as Jeffreys maintained, but represented subjective belief, and consequently one set of prior probabilities was as good as any other.

## 1.5. SYNOPSIS AND AIMS

The study starts in Chapter 2 with an account of the history of probability from the mid-seventeenth to the end of the nineteenth centuries, stressing the relative positions of the epistemic and frequency interpretations. For most of the eighteenth century they were in uneasy mixture, as the classical probabilists applied a single calculus to games of chance, annuities, the combination of observations, and rational action in the face of uncertainty. The distinction was only clearly drawn around the beginning of the nineteenth century. By 1850, the success of astronomical error theory, and the rise of social statistics – itself largely driven by bureaucrats to simplify taxation and administration – made the frequency interpretation more natural. The mathematics of the theory, however, remained intimately connected with particular implementations. Many new statistical techniques were developed by the biometric school expressly to attack particular problems of heredity and evolution, for example. (Ghosts of the biometricians' work survive in the current statistical terminology of regression and correlation.) Yet away from the reduction and analysis of data, many scientists continued to regard inverse probability, and the associated epistemic interpretation, as the model for uncertain inference.

Chapters 3 and 4 examine frequentist and Bayesian probability in more detail. I concentrate on the versions due to Fisher and Jeffreys, and argue that in each case, the choice of interpretation and its subsequent development evolved with a particular experience of scientific practice. Though the treatment is roughly symmetric, these chapters are of unequal length. With his major contributions both to statistics and the modern theory of evolution, Fisher is already recognized as a major figure in the history of science, and has generated a large secondary literature, including an excellent biography by his daughter, Joan Fisher Box. Jeffreys is less well known, yet deserves attention not only because of his contributions to geophysics and probability theory, but because his work style – the solitary theorist introducing rigorous mathematics to a range of fields previously associated with the vaguer methods of the enthusiast or dilettante – is now rare in a profession characterized by rigid boundaries between sub-disciplines. Moreover, though the frequency interpretation of probability and statistical model of experimental design is familiar from standard courses in statistics, the epistemic interpretation, despite the current revival of Bayesian probability theory, is still little known. Thus in addition to making the historiographic points – that even mathematical theories have no unique meaning, but are 'enculturated,' and therefore that a choice between them is not forced by content, but depends also on the context – my work is intended to serve as an introduction both to Bayesian probability theory and to the scientific career of Harold Jeffreys.

The backbone of the study, Chapter 5, concerns the Fisher–Jeffreys debate. By directly comparing frequentist and Bayesian methods, my discussion is intended to continue the explanations begun in previous chapters. It also makes the point that the difference between the approaches of Fisher and Jeffreys stemmed not from mere error or misunderstanding, but from distinct views of the character and purpose of scientific inquiry.

In Chapter 6 I move from Fisher and Jeffreys to ask why the Bayesian position was generally deserted in the years up to the World War II. The discipline of statistics, like economics, effaces its history. Perhaps anxious to encourage the appearance of objectivity, statistics textbooks – especially primers for researchers in other fields – make little mention of controversies in the development of the subject. Thus Bayesianism is usually suppressed, and an inconsistent mixture of frequentist methods, taken from both the Fisherian and Neyman–Pearson schools, is presented as a harmonious 'statistics.' This section of the study is intended as a first step at correction. It outlines the development of probabilistic thought during the 1930s across the disciplines of statistics, the social sciences, the physical and biological sciences, mathematics, and philosophy.

Chapter 6 is not, however, intended to be comprehensive. I do little to distinguish the several variations on the frequentist theme that either emerged or were developed during the 1930s. These include the theories associated with the names of Neyman and Pearson, von Mises, Reichenbach, Popper, Kolmogorov, and Koopman. Each enjoyed a period of support in at least one field, but apart from the Neyman–Pearson theory, none exerted much general influence in the English-speaking scientific world. It is true that von Mises's approach had an impact, and that Kolmogorov's formulation, in terms of measure theory, laid the foundation for a fully consistent theory of probability. But the influence of both developments was largely limited to pure mathematicians rather than practicing scientists. Neither von Mises nor Kolmogorov was widely cited in statistics texts or debates of the Royal Statistical Society during the period I consider. Fisher might have been expected to welcome such frequentist contributions, yet rarely referred to their work in his papers, even when discussing the theoretical and mathematical foundations of the subject. He considered von Mises's theory too restrictive to be of much use in scientific research, and Kolmogorov's too mathematical.[16] Hence my discussions of von Mises and Kolmogorov in Chapter 6 are brief. For similar reasons, I do not consider the subjective theory of de Finetti here at all. Though he started work on probability in the late 1920s, and developed his theory in a number of important papers through the 1930s, de Finetti was an isolated figure until the 1950s even in Italy. Jeffreys saw none of his work, and had not even heard of him until around 1983.[17] The main aim of Chapter 6 is not to distinguish these various sub-categories, but to chart the growing ascendancy of frequentistism over Bayesianism during the 1930s.

The final Chapter 7 is in the form of a brief epilogue followed by conclusions, and shows how the methodological position outlined in this introduction might be applied to the post-war history of probability.

Appendix 1 is a short bibliographical essay of sources used in Chapter 2. Appendix 2 presents the mathematical case for inverse probability as a model of scientific inference. Appendix 3 lists the abbreviations used in the footnotes.

[16] In a letter of 1946 Fisher ironically reported his discovery that some of his arguments had apparently been justified by "a certain, I believe very learned, Russian named Kolmogoroff"; he was still claiming to be unfamiliar with Kolmogorov's axioms around 1950.

[17] For more on von Mises, Kolmogorov, and de Finetti, see von Plato 1994.